



(19)

Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 874 045 A1

(12)

EUROPEAN PATENT APPLICATION

published in accordance with Art. 158(3) EPC

(43) Date of publication:

28.10.1998 Bulletin 1998/44

(51) Int. Cl.⁶: **C12N 15/00, C12P 21/00**

(21) Application number: 97935810.8

(86) International application number:
PCT/JP97/02859

(22) Date of filing: 19.08.1997

(87) International publication number:
WO 98/07840 (26.02.1998 Gazette 1998/08)

(84) Designated Contracting States:

AT BE CH DE DK ES FR GB IE IT LI LU NL SE

(30) Priority: 19.08.1996 JP 235928/96

(83) Declaration under Rule 28(4) EPC (expert solution)

(71) Applicant:

SNOW BRAND MILK PRODUCTS CO., LTD.
Sapporo-shi, Hokkaido 065 (JP)

(72) Inventors:

- NAKAGAWA, Nobuaki,
Nishlura Heights 2-4
Shimotsuga-gun, Tochigi 329-05 (JP)
- YASUDA, Hisataka
Kawachi-gun, Tochigi 329-04 (JP)
- MORINAGA, Tomonori
Shimotsuga-gun, Tochigi 321-02 (JP)

(74) Representative:

Wakerley, Helen Rachael
Reddie & Grose,
16 Theobalds Road
London WC1X 8PL (GB)

(54) NOVEL DNAs AND PROCESS FOR PRODUCING PROTEINS BY USING THE SAME

(57) DNAs having the nucleotide sequences of the Sequences No. 1 and No. 2 in the Sequence Table and a process for producing a protein which comprises inserting these DNAs into expression vectors to thereby produce a protein having molecular weights of about 60 kD (under reductive conditions) and about 60 kD and 120 kD (under non-reductive conditions) and being capable of inhibiting formation of osteoclast. These proteins are useful in the treatment of osteoporosis and rheumatism.

Description**FIELD OF TECHNOLOGY**

5 The present invention relates to a novel DNA and a process for preparing a protein which possesses an activity to inhibit osteoclast differentiation and/or maturation (hereinafter called osteoclastogenesis-inhibitory activity) by a genetic engineering technique using the DNA. More particularly, the present invention relates to a genomic DNA encoding a protein OCIF which possesses an osteoclastogenesis-inhibitory activity and a process for preparing said protein by a genetic engineering technique using the genomic DNA.

BACKGROUND OF THE INVENTION

10 Human bones are constantly repeating a process of resorption and formation. Osteoblasts controlling formation of bones and osteoclasts controlling resorption of bones take major roles in this process. Osteoporosis is a typical disease caused by abnormal metabolism of bones. This disease is caused when bone resorption by osteoclasts exceeds bone formation by osteoblasts. Although the mechanism of this disease is still to be elucidated completely, the disease causes the bones to ache, makes the bones fragile, and may result in fracturing of the bones. As the population of the aged increases, this disease results in an increase in bedridden aged people which becomes a social problem. Urgent development of a therapeutic agent for this disease is strongly desired. Disease due to a decrease in bone mass is 15 expected to be treated by controlling bone resorption, accelerating bone formation, or improving balance between bone resorption and formation.

20 Osteogenesis is expected to increase by accelerating proliferation, differentiation, or activation of the cells controlling bone formation, or by controlling proliferation, differentiation, or activation of the cells involved in bone resorption. In recent years, strong interest has been directed to physiologically active proteins (cytokines) exhibiting such activities 25 as described above, and energetic research is ongoing on this subject. The cytokines which have been reported to accelerate proliferation or differentiation of osteoblasts include the proteins of fibroblast growth factor family (FGF: Rodan S. B. et al., Endocrinology vol. 121, p 1917, 1987), insulin-like growth factor I (IGF-I: Hock J. M. et al., Endocrinology vol. 122, p 254, 1988), insulin growth factor II (IGF-II: McCarthy T. et al., Endocrinology vol. 124, p 301, 1989), Activin A (Centrella M. et al., Mol. Cell. Biol., vol. 11, p 250, 1991), transforming growth factor- β , (Noda M., The Bone, vol. 2, p 29, 1988), Vasculotropin (Varonique M. et al., Biochem. Biophys. Res. Commun., vol. 199, p 380, 1994), and the protein of heterotopic bone formation factor family (bone morphogenic protein; BMP: BMP-2; Yanaguchi A. et al., J. Cell Biol. vol. 113, p 682, 1991, OP-1; Sampath T. K. et al., J. Biol. Chem. vol. 267, p 20532, 1992, and Knutsen R. et al., Biochem. Biophys. Res. Commun. vol. 194, p 1352, 1993).

30 On the other hand, as the cytokines which suppress differentiation and/or maturation of osteoclasts, transforming growth factor- β (Chenu C. et al., Proc. Natl. Acad. Sci. USA, vol. 85, p 5683, 1988), interleukin-4 (Kasano K. et al., Bone-Miner., vol. 21, p 179, 1993), and the like have been reported. Further, as the cytokines which suppress bone resorption by osteoclast, calcitonin (Bone-Miner., vol. 17, p 347, 1992), macrophage colony stimulating factor (Hattersley G. et al., J. Cell. Physiol. vol. 137, p 199, 1988), interleukin-4 (Watanabe, K. et al., Biochem. Biophys. Res. Commun. vol. 172, p 1035, 1990), and interferon- γ (Gowen M. et al., J. Bone Miner. Res., vol. 1, p 46.9, 1986) have been 35 reported.

40 These cytokines are expected to be used as agents for treating diseases accompanying bone loss by accelerating bone formation or suppressing of bone resorption. Clinical tests are being undertaken to verify the effect of improving bone metabolism of some cytokines such as insulin-like growth factor-I and the heterotopic bone formation factor family. In addition, calcitonin is already commercially available as a therapeutic agent for osteoporosis and a pain relief agent. 45 At present, drugs for clinically treating bone diseases or shortening the period of treatment of bone diseases include activated vitamin D₃, calcitonin and its derivatives, and hormone preparations such as estradiol agent, ipriflavon or calcium preparations. These agents are not necessarily satisfactory in terms of the efficacy and therapeutic results. Development of a novel therapeutic agent which can be used in place of these agents is strongly desired.

45 In view of this situation, the present inventors have undertaken extensive studies. As a result, the present inventors 50 had found protein OCIF exhibiting an osteoclastogenesis-inhibitory activity in a culture broth of human embryonic lung fibroblast IMR-90 (ATCC Deposition No. CCL186), and filed a patent application (PCT/JP96/00374). The present inventors have conducted further studies relating to the origin of this protein OCIF exhibiting the osteoclastogenesis-inhibitory activity. The studies have matured into determination of the sequence of a genomic DNA encoding the human origin OCIF. Accordingly, an object of the present invention is to provide a genomic DNA encoding protein OCIF exhibiting 55 osteoclastogenesis-inhibitory activity and a process for preparing this protein by a genetic engineering technique using the genomic DNA.

DISCLOSURE OF THE INVENTION

Specifically, the present invention relates to a genomic DNA encoding protein OCIF exhibiting osteoclastogenesis-inhibitory activity and a process for preparing this protein by a genetic engineering technique using the genomic DNA.

5 The DNA of the present invention includes the nucleotide sequences No. 1 and No. 2 in the Sequence Table attached hereto.

Moreover, the present invention relates to a process for preparing a protein, comprising inserting a DNA including the nucleotide sequences of the sequences No. 1 and No. 2 in the Sequence Table into an expression vector, producing a vector capable of expressing a protein having the following physicochemical characteristics and exhibiting the activity 10 of inhibiting differentiation and/or maturation of osteoclasts, and producing this protein by a genetic engineering technique,

(a) molecular weight (SDS-PAGE):

15 (i) Under reducing conditions: about 60 kD,
 (ii) Under non-reducing conditions: about 60 kD and about 120 kD;

(b) amino acid sequence:

includes an amino acid sequence of the Sequence ID No. 3 of the Sequence Table,

20 (c) affinity:
 exhibits affinity to a cation exchanger and heparin, and
 (d) thermal stability:

25 (i) the osteoclast differentiation and/or maturation inhibitory activity is reduced when treated with heat at 70°C for 10 minutes or at 56°C for 30 minutes,
 (ii) the osteoclast differentiation and/or maturation inhibitory activity is lost when treated with heat at 90°C for 10 minutes.

30 The protein obtained by expressing the gene of the present invention exhibits an osteoclastogenesis-inhibitory activity. This protein is effective as an agent for the treatment and improvement of diseases involving decrease in the amount of bone such as osteoporosis, diseases relating to bone metabolism abnormality such as rheumatism, degenerative joint disease, or multiple myeloma, and is useful as an antigen to establish an immunological diagnosis of such diseases.

35 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a result of Western Blotting analysis of the protein obtained by causing genomic DNA of the present invention to express a protein in Example 4 (iii), wherein lane 1 indicates a marker, lane 2 indicates the culture broth of COS7 cells in which a vector pWESR α OCIF (Example 4 (iii)) has been transfected, and lane 3 is the culture broth of 40 COS7 cell in which a vector pWESR α (control) has been transfected.

BEST MODE FOR CARRYING OUT THE INVENTION

45 The genomic DNA encoding the protein OCIF which exhibits osteoclastogenesis-inhibitory activity in the present invention can be obtained by preparing a cosmid library using a human placenta genomic DNA and a cosmid vector and by screening this library using DNA fragments which are prepared based on the OCIF cDNA as a probe. The thus-obtained genomic DNA is inserted into a suitable expression vector to prepare an OCIF expression cosmid. A recombinant type OCIF can be obtained by transfecting the genomic DNA into a host organism such as various types of cells or microorganism strains and causing the DNA to express a protein by a conventional method. The resultant protein 50 exhibiting osteoclastogenesis-inhibitory activity (an osteoclastogenesis-inhibitory factor) is useful as an agent for the treatment and improvement of diseases involving a decrease in bone mass such as osteoporosis and other diseases relating to bone metabolism abnormality and also as an antigen to prepare antibodies for establishing immunological diagnosis of such diseases. The protein of the present invention can be prepared as a drug composition for oral or non-oral administration. Specifically, the drug composition of the present invention containing the protein which is an osteoclastogenesis-inhibitory factor as an active ingredient can be safely administered to humans and animals. As the form 55 of drug composition, a composition for injection, composition for intravenous drip, suppository, nasal agent, sublingual agent, percutaneous absorption agent, and the like are given. In the case of the composition for injection, such a composition is a mixture of a pharmacologically effective amount of osteoclastogenesis-inhibitory factor of the present

invention and a pharmaceutically acceptable carrier. The composition may further comprise amino acids, saccharides, cellulose derivatives, and other excipients and/or activation agents, including other organic compounds and inorganic compounds which are commonly added to a composition for injection. When an injection preparation is prepared using the osteoclastogenesis-inhibitory factor of the present invention and these excipients and activation agents, a pH adjuster, buffering agent, stabilizer, solubilizing agent, and the like may be added if necessary to prepare various types of injection agents.

The present invention will now be described in more detail by way of examples which are given for the purpose of illustration and not intended to be limiting of the present invention.

10 **Example 1**

(Preparation of a cosmid library)

15 A cosmid library was prepared using human placenta genomic DNA (Clonetech; Cat. No. 6550-2) and pWE15 cosmid vector (Stratagene). The experiment was carried out following principally the protocol attached to the pWE15 cosmid vector kit of Stratagene Company, provided Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory (1989)) was referred to for common procedures for handling DNA, *E. coli*, and phage.

20 (i) Preparation of restrictive enzymolysate of human-genomic DNA

25 Human placenta genomic DNA dissolved in 750 μ l of a solution containing 10 mM Tris-HCl, 10 mM MgCl₂, and 100 mM NaCl was added to four 1.5 ml Eppendorf tubes (tube A, B, C, and D) in the amount of 100 μ g each. Restriction enzyme MboI was added to these tubes in the amounts of 0.2 unit for tube A, 0.4 unit for tube B, 0.6 unit for tube C, and 0.8 unit for tube D, and DNA was digested for 1 hour. Then, EDTA in the amount to make a 20 mM concentration was added to each tube to terminate the reaction, followed by extraction with phenol/chloroform (1:1). A two-fold amount of ethanol was added to the aqueous layer to precipitate DNA. DNA was collected by centrifugation, washed with 70% ethanol, and DNA in each tube was dissolved in 100 μ l of TE (10 mM HCl (pH 8.0) + 1 mM EDTA buffer solution, hereinafter called TE). DNA in four tubes was combined in one tube and incubated for 10 minutes at 68°C. After cooling to room temperature, the mixture was overlayed onto a 10%-40 % linear sucrose gradient which was prepared in a buffer containing 20 mM Tris-HCl (pH 8.0), 5 mM EDTA, and 1 mM NaCl in an centrifugal tube (38 ml). The tube was centrifuged at 26,000 rpm for 24 hours at 20°C using a rotor SRP28SA manufactured by Hitachi, Ltd. and 0.4 ml fractions of the sucrose gradient was collected using a fraction collector. A portion of each fraction was subjected to 0.4% agarose electrophoresis to confirm the size of DNA. Fractions containing DNA with a length of 30 kb (kilo base pair) to 40 kb were thus combined. The DNA solution was diluted with TE to make a sucrose concentration to 10% or less and 2.5-fold volumes of ethanol was added to precipitate DNA. DNA was dissolved in TE and stored at 4°C.

30 (ii) Preparation of cosmid vector

35 The pWE15 cosmid vector obtained from Stratagene Company was completely digested with restriction enzyme BamHI according to the protocol attached to the cosmid vector kit. DNA collected by ethanol precipitation was dissolved in TE to a concentration of 1 mg/ml. Phosphoric acid at the 5'-end of this DNA was removed using calf small intestine alkaline phosphatase, and DNA was collected by phenol extraction and ethanol precipitation. The DNA was dissolved in TE to a concentration of 1 mg/ml.

40 (iii) Ligation of genomic DNA to vector and in vitro packaging

45 1.5 micrograms of genomic DNA fractionated according to size and 3 μ g of pWE15 cosmid vector which was digested with restriction enzyme BamHI were ligated in 20 μ l of a reaction solution using Ready-To-Go T4DNA ligase of Pharmacia Company. The ligated DNA was packaged in vitro using Gigapack™ II packaging extract (Stratagene) according to the protocol. After the packaging reaction, a portion of the reaction mixture was diluted stepwise with an SM buffer solution and mixed with *E. coli* XL1-Blue MR (Stratagene) which was suspended in 10 mM MgCl₂ to cause phage to infect, and plated onto LB agar plates containing 50 μ g/ml of ampicillin. The number of colonies produced was counted. The number of colonies per 1 μ l of packaging reaction was calculated based on this result.

50 (iv) Preparation of a cosmid library

55 The packaging reaction solution thus prepared was mixed with *E. coli* XL1-Blue MR and the mixture was plated onto agarose plates containing ampicillin so as to produce 50,000 colonies per agarose plate having a 15 cm of diam-

eter. After incubating the plate overnight at 37°C, an LB culture medium was added in the amount of 3 ml per plate to suspend and collect colonies of *E. coli*. Each agarose plate was again washed with 3 ml of the LB culture medium and the washing was combined with the original suspension of *E. coli*. The *E. coli* collected from all agarose plates was placed in a centrifugal tube, glycerol was added to a concentration of 20%, and ampicillin was further added to make a final concentration of 50 µg/ml. A portion of the *E. coli* suspension was removed and the remainder was stored at -80°C. The removed *E. coli* was diluted stepwise and plated onto an agar plates to count the number of colonies per 1 ml of suspension.

5 **Example 2**

10

(Screening of cosmid library and purification of colony)

A nitrocellulose filter (Millipore) with a diameter of 14.2 cm was placed on each LB agarose plate with a diameter of 15 cm which contained 50 µg/ml of ampicillin. The cosmid library was plated onto the plates so as to produce 50,000 15 colonies of *E. coli* per plate, followed by incubation overnight at 37°C. *E. coli* on the nitrocellulose filter was transferred to another nitrocellulose filter according to a conventional method to obtain two replica filters. According to the protocol attached to the cosmid vector kit, cosmid DNA in the *E. coli* on the replica filters was denatured with an alkali, neutralized, and immobilized on the nitrocellulose filter using a Stratalinker (Stratagene). The filters were heated for two hours 20 at 80°C in a vacuum oven. The nitrocellulose filters thus obtained were hybridized using two kinds of DNA produced, respectively, from 5'-end and 3'-end of human OCIF cDNA as probes. Namely, a plasmid was purified from *E. coli* pKB/OIF10 (deposited at The Ministry of International Trade and Industry, the Agency of Industrial Science and Technology, Biotechnology Laboratory, Deposition No. FERM BP-5267) containing OCIF cDNA. The plasmid containing OCIF cDNA was digested with restriction enzymes KpnI and EcoRI. Fragments thus obtained was separated using agarose gel electrophoresis. KpnI/EcoRI fragment with a length of 0.2 kb was purified using a QIAEX II gel extraction 25 kit (Qiagen). This DNA was labeled with ³²P using the Megaprime DNA Labeling System (Amasham) (5'-DNA probe). Apart from this, a BamHI/EcoRV fragment with a length of 0.2 kb which was produced from the above plasmid by digestion with restriction enzymes BamHI and EcoRV was purified and labeled with ³²P (3'-DNA probe). One of the replica filters described above was hybridized with the 5'-DNA probe and the other with the 3'-DNA probe. Hybridization and 30 washing of the filters were carried out according to the protocol attached to the cosmid vector kit. Autoradiography detected several positive signals with each probe. One colony which gave positive signals with both probe was identified. The colony on the agar plate, which corresponding to the signal on the autoradiogram was isolated and purified. A cosmid was prepared from the purified colony by a conventional method. This cosmid was named pWEOCIF. The size of human genomic DNA contained in this cosmid was about 38 kb.

35 **Example 3**

(Determination of the nucleotide sequence of human OCIF genomic DNA)

40 **(i) Subcloning of OCIF genomic DNA**

45

Cosmid pWEOCIF was digested with restriction enzyme EcoRI. After the separation of the DNA fragments thus produced by electrophoresis using a 0.7% agarose gel, the DNA fragments were transferred to a nylon membrane (Hybond -N, Amasham) by the Southern blot technique and immobilized on the nylon membrane using Stratalinker (Stratagene). On the other hand, plasmid pBKOCIF was digested with restriction enzyme EcoRI and a 1.6 kb fragment containing human OCIF cDNA was isolated by agarose gel electrophoresis. The fragment was labeled with ³²P using 50 the Megaprime DNA labeling system (Amasham).

Hybridization of the nylon membranes described above with the ³²P-labeled 1.6-kb OCIF cDNA was performed according to a conventional method detected that DNA fragments with a size of 6 kb, 4 kb, 3.6 kb, and 2.6 kb. These fragments hybridized with the human OCIF cDNA were isolated using agarose gel electrophoresis and individually subcloned into an EcoRI site of pBluescript II SK + vector (Stratagene) by a conventional method. The resulting plasmids were respectively named pBSE 6, pBSE 4, pBSE 3.6, and PBSE 2.6.

55 **(ii) Determination of the nucleotide sequence**

The nucleotide sequence of human OCIF genomic DNA which was subcloned into the plasmid was determined using the ABI Dideoxy Terminator Cycle Sequencing Ready Reaction kit (Perkin Elmer) and the 373 Sequencing System (Applied Biosystems). The primer used for the determination of the nucleotide sequence was synthesized based on the nucleotide sequence of human OCIF cDNA (Sequence ID No. 4 in the Sequence Table). The nucleotide

sequences thus determined are given as the Sequences No. 1 and No. 2 in the Sequence Table. The Sequence ID No. 1 includes the first exon of the OCIF gene and the Sequence ID No. 2 includes the second, third, fourth, and fifth exons. A stretch of about 17 kb is present between the first and second exons.

5 Example 4

(Production of recombinant OCIF using COS-7 cells)

(i) Preparation of OCIF genomic DNA expression cosmid

10 To express OCIF genomic DNA in animal cells, an expression unit of expression plasmid pcDL-SR α 296 (Molecular and Cellar Biology, vol. 8, P466-472, 1988) was inserted into cosmid vector pWE15 (Stratagene). First of all, the expression plasmid pcDL-SR α 296 was digested with a restriction enzyme Sal I to cut out expression unit with a length of about 1.7 kb which includes an SR α promotor, SV40 later splice signal, poly (A) addition signal, and so on. The digestion products were separated by agarose electrophoresis and the 1.7-kb fragment was purified using the QIAEX II gel extraction kit (Qiagen). On the other hand, cosmid vector pWE15 was digested with a restriction enzyme EcoRI and fragments were separated using agarose gel electrophoresis. pWE15 DNA of 8.2 kb long was purified using the QIAEX II gel extraction kit (Qiagen). The ends of these two DNA fragments were blunted using a DNA blunting kit (Takara Shuzo), ligated using a DNA ligation kit (Takara Shuzo), and transferred into E. coli DH5 α (Gibco BRL). The resultant transformant was grown and the expression cosmid pWESR α containing an expression unit was purified using a Qiagen column (Qiagen).

15 The cosmid pWE OCIF containing the OCIF genomic DNA with a length of about 38 kb obtained in (i) above was digested with a restriction enzyme NotI to cut out the OCIF genomic DNA of about 38 kb. After separation by agarose gel electrophoresis, the DNA was purified using the QIAEX II gel extraction kit (Qiagen). On the other hand, the expression cosmid pWESR α was digested with a restriction enzyme EcoRI and the digestion product was extracted with phenol and chloroform, ethanol-precipitated, and dissolved in TE.

20 pWESR α digested with a restriction enzyme EcoRI and an EcoRI-Xmnl-NotI adapter (#1105, #1156 New England Biolaboratory Co.) were ligated using T4 DNA ligase (Takara Shuzo Co., Ltd.). After removal of the free adapter by agarose gel electrophoresis, the product was purified using QIAEX gel extraction kit (Qiagen). The OCIF genomic DNA with a length of about 37 kb which was derived from the digestion with restriction enzyme NotI and the pWESR α to which the adapter was attached were ligated using T4 DNA ligase (Takara Shuzo). The DNA was packaged in vitro using the Gigapack packaging extract (Stratagene) and infected with E. coli XL1-Blue MR (Stratagene). The resultant transformant was grown and the expression cosmid pWESR α OCIF which contained OCIF genomic DNA was inserted was purified using a Qiagen column (Qiagen). The OCIF expression cosmid pWESR α OCIF was ethanol-precipitated and dissolved in sterile distilled water and used in the following analysis.

(ii) Transient expression of OCIF genomic DNA and measurement of OCIF activity

25 A recombinant OCIF was expressed as described below using the OCIF expression cosmid pWESR α OCIF obtained in (i) above and its activity was measured. COS-7 (8×10^5 cells/well) cells (Riken Cell Bank, RCB0539) were planted in a 6-well plate using DMEM culture medium (Gibco BRL) containing 10% fetal bovine serum (Gibco BRL). On the following day, the culture medium was removed and cells were washed with serum-free DMEM culture medium. The OCIF expression cosmid pWESR α OCIF which had been diluted with OPTI-MEM culture medium (Gibco BRL) was mixed with Lipopfectamine and the mixture was added to the cells in each well according to the attached protocol. The expression cosmid pWESR α was added to the cells in the same manner as a control. The amount of the cosmid DNA and Lipopfectamine was respectively 3 μ g and 12 μ l. After 24 hours, the culture medium was removed and 1.5 ml of fresh EX-CELL 301 culture medium (JRH Bioscience) was added to each well. The culture medium was recovered after 48 hours and used as a sample for the measurement of OCIF activity. The measurement of OCIF activity was carried out according to the method described by Kumegawa, M. et al. (Protein, Nucleic Acid, and Enzyme, Vol. 34, p 999 (1989)) and the method of TAKAHASHI, N. et al. (Endocrinology vol. 122, p 1373 (1988)). The osteoclast formation in the presence of activated vitamin D₃ from bone marrow cells isolated from mice aged about 17 days was evaluated by the induction of tartaric acid resistant acidic phosphatase activity. The inhibition of the acid phosphatase was measured and used as the activity of the protein which possesses osteoclastogenesis-inhibitory activity (OCIF). Namely, 100 μ l/well of a OCIF sample which was diluted with α -MEM culture medium (Gibco BRL) containing 2×10^{-8} M activated vitamin D₃ and 10% fetal bovine serum was added to each well of a 96 well micro plate. Then, 3×10^5 bone marrow cells isolated from mice (about 17-days old) suspended in 100 μ l of α -MEM culture medium containing 10% fetal bovine serum were added to each well of the 96 well micro plate and cultured for a week at 37°C and 100% humidity under 5% CO₂ atmosphere. On days 3 and 5, 160 μ l of the conditioned medium was removed from each well, and 160 μ l of a sam-

ple which was diluted with α -MEM culture medium containing 1×10^{-8} M activated vitamin D₃ and 10% fetal bovine serum was added. After 7 days from the start of culturing, the cells were washed with a phosphate buffered saline and fixed with a ethanol/acetone (1:1) solution for one minute at room temperature. The osteoclast formation was detected by staining the cells using an acidic phosphatase activity measurement kit (Acid Phosphatase, Leucocyte, Cat.No. 387-A, Sigma Company). A decrease in the number of cells positive to acidic phosphatase activity in the presence of tartaric acid was taken as the OCIF activity. The results are shown in Table 1, which indicates that the conditioned medium exhibits the similar activity to natural type OCIF obtained from the IMR-90 culture medium and recombinant OCIF produced by CHO cells.

10

TABLE 1

Activity of OCIF expressed by COS-7 cells in the conditioned medium						
Dilution	1/10	1/20	1/40	1/80	1/160	1/320
OCIF genomic DNA introduced	++	++	++	++	+	-
Vector introduced	-	-	-	-	-	-
Untreated	-	-	-	-	-	-

"++" indicates an activity inhibiting 80% or more of osteoclast formation, "+" indicates an activity inhibiting 30-80% of osteoclast formation, and "-" indicates that no inhibition of osteoclast formation is observed.

(iii) Identification of the product by Western Blotting

25 A buffer solution (10 μ l) for SDS-PAGE (0.5 M Tris-HCl, 20% glycerol, 4% SDS, 20 μ g/ml bromophenol blue, pH 6.8) was added to 10 μ l of the sample for the measurement of OCIF activity prepared in (ii) above. After boiling for 3 minutes at 100°C, the mixture was subjected to 10% SDS polyacrylamide electrophoresis under non-reducing conditions. The proteins were transferred from the gel to a PVDF membrane (ProBlott, Perkin Elmer) using semi-dry blotting apparatus (Biorad). The membrane was blocked and incubated for 2 hours at 37°C together with a horseradish peroxidase-labeled anti-OCIF antibody obtained by labeling the previously obtained OCIF protein with horseradish peroxidase according to a conventional method. After washing, the protein which has bound the anti-OCIF antibody was detected using the ECL system (Amasham). As shown in Figure 1, two bands, one with a molecular weight of about 120 kilo dalton and the other 60 kilo dalton, were detected in the supernatant obtained from the culture broth of COS-7 cells in which pWESR α OCIF was transfected. On the other hand, these two bands with a molecular weight of about 30 120 kilo dalton and 60 kilo dalton were not detected in the supernatant obtained from the culture broth of COS-7 cells in which pWESRvector was transfected, confirming that the protein obtained was OCIF.

INDUSTRIAL APPLICABILITY

40 The present invention provides a genomic DNA encoding a protein OCIF which possesses an osteoclastogenesis-inhibitory activity and a process for preparing this protein by a genetic engineering technique using the genomic DNA. The protein obtained by expressing the gene of the present invention exhibits an osteoclastogenesis-inhibitory activity and is useful as an agent for the treatment and improvement of diseases involving a decrease in the amount of bone such as osteoporosis, other diseases resulting from bone metabolism abnormality such as rheumatism or degenerative 45 joint disease, and multiple myeloma. The protein is further useful as an antigen to establish antibodies useful for an immunological diagnosis of such diseases.

NOTE ON MICROORGANISM

50 Depositing Organization:

The Ministry of International Trade and Industry, National Institute of Bioscience and Human Technology, Agency of Industrial Science and Technology

Address: 1-3, Higashi-1-Chome, Tsukuba-shi, Ibaraki-ken, Japan

Date of Deposition: June 21, 1995 (originally deposited on June 21, 1995 and transferred to the international deposition according to the Budapest Treaty on October 25, 1995)

55 Accession No. FERM BP-5267

TABLE OF SEQUENCES

5

Sequence number: 1

Length of sequence: 1316

10

Sequence Type: nucleic acid

Strandedness: double

15

Topology: linear

Molecular type: genomic DNA (human OCIF genomic DNA-1)

20

Sequence:

CTGGAGACAT	ATAACTTGAA	CACTTGGCCC	TGATGGGAA	GCAGCTCTGC	AGGGACTTTT	60
TCAGCCATCT	GTAAACAATT	TCAGTCCAA	CCCGCGAACT	GTAATCCATG	AATGGGACCA	120
CACTTTACAA	GTCATCAAGT	CTAACTTCTA	GACCAGGGAA	TTAATCCGGG	AGACAGCGAA	180
CCCTAGAGCA	AACTGCCAAA	CTTCTCTCGA	TAGCTTGAGG	CTAGTGGAAA	GACCTCGAGG	240
AGGCTACTCC	ACAAGTTCAAG	CGCGTAGGAA	GCTCCGATAC	CAATAGCCCT	TTGATGATGG	300
TGGGGTTGGT	GAAGGGAAACA	GTGCTCCGCA	AGGTTATCCC	TGCCCCAGGC	AGTCCAATTT	360
TCACTCTGCA	GATTCTCTCT	GGCTCTAACT	ACCCCGATA	ACAAGGAGTG	AATCCAGAAT	420
AGCACGGGCT	TTAGGGCCAA	TCAGACATTA	CTTAGAAAAA	TTCTCTACTAC	ATGGTTTATG	480
TAAACTTGAA	GATGAATCAT	TGCGAACTCC	CCGAAAAGGG	CTCAGACAAAT	GCCATGCCATA	540
AAGAGGGGCC	CTGTAATTG	AGGTTTCAGA	ACCCGAAGTG	AAGGGGTCA	GCAGCCGGGT	600
ACGGCCGAAA	CTCACAGCTT	TCGCCAGCG	AGAGGACAAA	GGTCTGGGAC	ACACTCCAAC	660
TGCGTCCGGA	TCTTGGCTGG	ATCGGACTCT	CAGGGTGGAG	GAGACACAAG	CACAGCAGCT	720
GCCCCAGCGTG	TGCCCAGCCC	TCCCACCGCT	GGTCCCGGCT	GCCAGGAGGC	TGGCCGCTGG	780
CGGGAAGGGG	CCGGGAAACC	TCAGAGCCCC	GGGGAGACAG	CAGCCGCCCT	GTTCTCAGC	840
CCGGTGGCTT	TTTTTCCCC	TGCTCTCCCA	GGGGACAGAC	ACCACCGCCC	CACCCCTCAC	900
GCCCCACCTC	CCTGGGGGAT	CCTTCCCCC	CCAGCCCTGA	AAGCGTTAAT	CCTGGAGCTT	960
TCTGCACACC	CCCCGACCCG	TCCCCCCC	GCTTCCTAAA	AAAGAAAGGT	GCAAAGTTG	1020
GTCCAGGATA	AAAAAATCAC	TGATCAAAGG	CAGGCAGATA	TTCTGTTGC	GGGGACGCTA	1080
TATATAACCT	GATGAGCCCA	CGGGCTGGGG	AGACCCACCG	GAGCGCTCCG	CCACCCCCCG	1140

55

CCTCCAAGCC CCTGAGGTTT CCGGGGACCA CA ATG AAC AAG TTC CTG TCC TGC 1193

5 Met Asn Lys Leu Leu Cys Cys
-20 -15

10 GCG CTC GTG GTAACTCCCT GGGCCAGCCC ACGGGTGCCTGGG 1242

Ala Leu Val

15 GAGGCTGCTG CCACCTGGTC TCCCCAACCTC CCAGGGGACCC GGCGGGGAGA AGGCTCCACT 1302
CGCTCCCTCC CAGG 1316

20 Sequence number: 2

Length of sequence: 9898

25 Sequence Type: nucleic acid

Strandedness: double

30 Topology: linear

Molecular type: genomic DNA (human OCIF genomic DNA-2)

35 Sequence:

GCTTACTTTG TGCCAAATCT CATTAGGCTT AAGGTAATAC AGGACTTTGA GTCAAATGAT 60
ACTGTTGCAC ATAAGAACAA ACCTATTTTC ATGCTAAGAT GATGCCACTG TGTTCCCTTC 120
40 TCCTTCTAG TTT CTG GAC ATC TCC ATT AAG TGG ACC ACC CAG GAA ACG TTT 171

Phe Leu Asp Ile Ser Ile Lys Trp Thr Thr Gln Glu Thr Phe

45 -10 -5 1

50 CCT CCA AAG TAC CTT CAT TAT GAC GAA GAA ACC TCT CAT CAG CTG TTG 219
Pro Pro Lys Tyr Leu His Tyr Asp Glu Glu Thr Ser His Gln Leu Leu

5 10 15

55

5	TGT GAC AAA TGT CCT CCT GGT ACC TAC CTA AAA CAA CAC TGT ACA GCA	267
	Cys Asp Lys Cys Pro Pro Gly Thr Tyr Leu Lys Glu His Cys Thr Ala	
	20 25 30 35	
10	AAG TGG AAG ACC GTG TGC GCC CCT TGC CCT GAC CAC TAC TAC ACA GAC	315
	Lys Trp Lys Thr Val Cys Ala Pro Cys Pro Asp His Tyr Tyr Thr Asp	
15	40 45 50	
20	AGC TGG CAC ACC AGT GAC GAG TGT CTA TAC TGC ACC CCC GTG TGC AAG	363
	Ser Trp His Thr Ser Asp Glu Cys Leu Tyr Cys Ser Pro Val Cys Lys	
	55 60 65	
25	GAC CTG CAG TAC GTC AAG CAG GAG TGC AAT CGC ACC CAC AAC CGC GTG	411
	Glu Leu Glu Tyr Val Lys Gln Glu Cys Asn Arg Thr His Asn Arg Val	
30	70 75 80	
35	TGC GAA TGC AAG GAA GGG CGC TAC CTT GAG ATA GAG TTC TGC TTG AAA	459
	Cys Glu Cys Lys Glu Gly Arg Tyr Leu Glu Ile Glu Phe Cys Leu Lys	
40	85 90 95	
45	CAT AGG AGC TGC CCT CCT CGA TTT GGA GTG GTG CAA GCT G GTACGTGTCA	509
	His Arg Ser Cys Pro Pro Gly Phe Gly Val Val Gln Ala	
	100 105 110	
50	ATGTGCAGCA AAATTAATTA GGATCATGCA AAGTCAGATA GTTGTGACAG TTTAGGAGAA	569

CACTTTGTT CTGATGACAT TATAGGATAG CAAATTGCAA AGGTAATGAA ACCTGCCAGG 629
 5 TAGGTACTAT GTGTCTGGAG TGCTTCCAAA GGACCATTGC TCAGAGGAAT ACTTGCCAC 689
 TACAGGGCAA TTTAATGACA AATCTCAAAT GCAGCAAATT ATTCTCTCAT GAGATGCATG 749
 10 ATGGTTTTT TTTTTTTT TAAAGAAACA AACTCAACTT CCACTATTCA TAGTTGATCT 809
 ATACCTCTAT ATTCACTTC AGCATGGACA CCTTCAAACG GCAGCACTT TTGACAAACA 869
 TCAGAAATGT TAATTTATAC CAAGAGAGTA ATTATGCTCA TATTAATGAG ACTCTGGAGT 929
 15 GCTAACAAATA AGCAGTTATA ATTAATTATG TAAAAAATGA GAATGGTCAG CGGAATTGCA 989
 TTTCATTATT AAAAACAAAGG CTAGTTCTTC CTTTACCATG GGAGCTGAGT GTTGGGAGG 1049
 GTAAGGACTA TAGCAGAACATC TCTTCAATGA GCTTATTCTT TATCTTAGAC AAAACAGATT 1109
 20 CTCAAGCCAA GAGCAAGCAC TTGCCTATAA ACCAAGTGCT TTCTCTTTG CATTITGAAC 1169
 AGCATTGGTC AGGGCTCATG TGTATTGAAT CTTTAAACC AGTAACCCAC GTTTTTTTC 1229
 TGCCACATTT GCGAAGCTTC AGTGCAGCCT ATAACCTTTC ATAGCTTGAG AAAATTAAGA 1289
 25 GTATCCACTT ACTTAGATGG AAGAAGTAAT CAGTATAGAT TCTGATGACT CAGTTGAAG 1349
 CAGTGTTCCT CAACTGAAGC CCTGCTGATA TTTAAGAAA TATCTGGATT CCTAGGCTGG 1409
 30 ACTCCTTTT GTGGGCAGCT GTCCCTGCGCA TTGTAGAATT TTGGCAGCAC CCCTGGACTC 1469
 TAGCCACTAG ATACCAATAG CAGTCCTTCC CCCATGTGAC AGCCAAAAAT GTCTTCAGAC 1529
 ACTGTCAAAT GTGCCAGGT GGCAAAATCA CTCCCTGGTTG AGAACAGGGT CATCAATGCT 1589
 35 AAGTATCTGT AACTATTTA ACTCTAAAAA CTGTGATAT ACAAAAGTCTA AATTATTAGA 1649
 CGACCAATAAC TTTAGGTTA AAGGCATACA AATGAAACAT TCAAAATCA AAATCTATTC 1709
 40 TGTTTCTCAA ATAGTGAATC TTATAAAATT AATCACAGAA GATGCAAATT GCATCAGACT 1769
 CCCTTAAAT TCCTCTTCTGT ATGAGTATTG GAGGGAGGAA TTGGTGATAG TTCTACTTT 1829
 CTATTGGATG GTACTTTGAG ACTCAAAAGC TAAGCTAAGT TGTGTGTGTG TCAGGGTGG 1889
 45 GGGTGTGGAA TCCCATCAGA TAAAAGCAA TCCATGTAAT TCATTCACTA AGTTGTATAT 1949
 GTACAAAAAT GAAAAGTGGG CTATGCAGCT TGGAAACTAG AGAATTITGA AAAATAATGG 2009
 50 AAATCACAAG GATCTTCCTT AAATAAGTAA GAAAATCTGT TTGTAGAATG AAGCAAGCAG 2069
 GCAGCCAGAA GACTCAGAAC AAAAGTACAC ATTTACTCT GTGTACACTG GCAGCACAGT 2129

5 GGGATTTATT TACCTCTCCC TCCCTAAAAA CCCACACAGC GCTTCCTCTT GGGAAATAAG 2189
 AGGTTCCAG CCCAAAGAGA AGGAAAGACT ATGTGGTGT ACTCTAAAAA CTATTTAATA 2249
 ACCGTTTGT TGTTCCTGTT GCTGTTTGA AATCAGATTG TCTCCTCTCC ATATTTTATT 2309
 10 TACTTCATTC TGTAAATTCC TGTGGAATTAA CTTAGAGCAA GCATGCTGAA TTCTCAACTG 2369
 TAAAGCCAAA TTTCTCCATC ATTATAATT CACATTTGC CTGGCAGGTT ATAATTTTA 2429
 TATTTCCACT GATAGTAATA AGGTAAAATC ATTACTTAGA TGGATAGATC TTTTCATAA 2489
 15 AAAGTACCAT CAGTTATAGA GGGAAAGTCAT GTTCATGTT AGGAAGGTCA TTAGATAAAG 2549
 CTTCTGAATA TATTATGAAA CATTACTTCT GTCAATTCTTA GATTCTTTT GTAAATAAC 2609
 TTTAAAAGCT AACTTACCTA AAAGAAATAT CTGACACATA TGAACCTCTC ATTAGGATGC 2669
 20 AGGAGAACAC CCAAGCCACA GATATGTATC TGAAGAATGA ACAAGATTCT TAGGCCCGGC 2729
 ACGGTGGCTC ACATCTGTA TCTCAAGAGT TTGAGAGGTC AAGGCGGGCA GATCACCTGA 2789
 GGTCAAGGAGT TCAAGACCAG CCTGGCCAAC ATGATGAAAC CCTGCCTCTA CTAAAATAC 2849
 AAAAATTAGC AGGGCATGGT GGTGCATGCC TCCAACCCCTA GCTACTCAGG AGGCTGAGAC 2909
 25 AGGAGAATCT CTTGAACCCCT CGAGGCGGAG GTTGTGGTGA GCTGAGATCC CTCTACTGCA 2969
 CTCCAGCCTG GGTGACAGAG ATGAGACTCC GTCCCTGCCG CGGCCCCCGC CTTCCCCCCC 3029
 AAAAAGATTG TTCTTCATGC AGAACATACCG GCAGTCAACA AAGGGAGACC TGGGTCCAGG 3089
 TGTCCAAGTC ACTTATTTCG AGTAAATTAG CAATGAAAGA ATGCCATGGA ATCCCTCCCC 3149
 30 AAATACCTCT GCTTATGATA TTGTAGAATT TGATATAGAC TTGTATCCCA TTAAAGGAGT 3209
 AGGATGTAGT AGGAAAGTAC TAAAAACAAA CACACAAACA GAAAACCCCTC TTGCTTGT 3269
 35 AAGGTGGTTC CTAAGATAAT GTCAGTCAA TGCTGGAAAT AATATTTAAT ATGTCAAGGT 3329
 TTTAGGCTGT GTTTCCCT CCTGTTCTT TTTCTGCCA GCCCTTGTC ATTTTGCGAG 3389
 GTCAATGAAT CATGTAGAAA GAGACAGGAG ATGAAACTAG AACCACTCCA TTTGCCCT 3449
 40 TTTTTATTT TCTGGTTTG GTAAAAGATA CAATGAGGTA GGAGGTTGAG ATTTATAAAT 3509
 GAAGTTAAC AAGTTCTGT AGCTTGATT TTTCTCTTC ATATTGTTA TCTTGCATAA 3569
 45 GCCAGAATTG GCCTGTAAAA TCTACATATG GATATTGAAG TCTAAATCTG TTCAACTAGC 3629
 TTACACTAGA TGGAGATATT TTCATATTCA GATACACTGG AATGTATGAT CTAGCCATGC 3689

5 GTAATATAGT CAAGTGTG AAGGTATTTA TTTTTAATAG CGTCTTACT TGTGGACTGG 3749
 TTCAAGTTT TCTGCCAATG ATTTCTCAA ATTTATCAA TATTTTCCA TCATGAACTA 3809
 AAATGCCCTT GCAGTCACCC TTCTGAAGT TTGAACGACT CTGCTGTTT AAACAGTTA 3869
 AGCAAATGGT ATATCATCTT CCGTTACTA TGTAGCTTAA CTGCAGGCTT ACGCTTTGA 3929
 10 GTCAGCGGCC AACTTTATTG CCACCTCAA AAGTTATTA TAATGTTGTA AATTTTACT 3989
 TCTCAAGGTT ACCATACTTA GGAGTTGCTT CACAATTAGG ATTCAAGGAAA GAAAGAACCT 4049
 15 CAGTAGGAAC TCATTGGAAT TTAATGATGC AGCATTCAAT GCGTACTAAT TTCAAAGAAT 4109
 GATATTACAG CAGACACACA GCAGTTATCT TGATTTCTA GGAATAATTG TATGAAGAAT 4169
 ATGGCTGACA ACACGGCCTT ACTGCCACTC AGCGGAGGCT GGACTAATGA ACACCCTACC 4229
 20 CTTCTTCCT TTCTCTCAC ATTTCATGAG CGTTTGTAG GTAACGAGAA AATTGACTTC 4289
 CATTGCATT ACAAGGAGGA GAAACTGGCA AAGGGGATGA TGGTCCAAGT TTTGTCTGT 4349
 CTAATGAAGT GAAAAATGAA AATGCTAGAG TTTTGTGCAA CATAATAGTA GCAGTAAAAA 4409
 25 CCAAGTGAAA ACTCTTCCA AAACGTGTT AAGAGGGCAT CTGCTGGAA ACGATTTGAG 4469
 GAGAAGGTAC TAAATTGCTT GGTATTTCC GTAG GA ACC CCA GAG CGA AAT ACA 4523

30 Gly Thr Pro Glu Arg Asn Thr

115

35 GTT TGC AAA AGA TGT CCA GAT GGG TTC TTC TCA AAT GAG ACG TCA TCT 4571
 Val Cys Lys Arg Cys Pro Asp Gly Phe Phe Ser Asn Glu Thr Ser Ser
 40 120 125 130 135

45 AAA CCA CCC TGT AGA AAA CAC ACA AAT TGC AGT GTC TTT GGT CTC CTG 4619
 Lys Ala Pro Cys Arg Lys His Thr Asn Cys Ser Val Phe Gly Leu Leu
 50 140 145 150

CTA ACT CAG AAA GGA AAT GCA ACA CAC GAC AAC ATA TGT TCC GGA AAC 4667

55

Leu Thr Gln Lys Gly Asn Ala Thr His Asp Asn Ile Cys Ser Gly Asn

5 155 160 165

10 ACT GAA TCA ACT CAA AAA TGT GGA ATA G GTAATTACAT TCCAAAATAC 4715

Ser Glu Ser Thr Gln Lys Cys Gly Ile

15 170 175

15 GTCTTTGTAC GATTTTGTAG TATCATCTCT CTCTCTGAGT TGAACACAAG GCCTCCACCC 4775

ACATTCTTGG TCAAACCTTAC ATTTTCCCTT TCTTGAATCT TAACCAGCTA AGGCTACTCT 4835

20 CGATGCATTA CTGCTAAAGC TACCACTCAG AATCTCTCAA AAACTCATCT TCTCACAGAT 4895

AACACCTCAA AGCTTCATTT TCTCTCCTTT CACACTGAAA TCAAATCTTG CCCATAGGCA 4955

AAGGGCAGTG TCAAGTTTGC CACTGAGATG AAATTAGGAG AGTCCAAACT GTAGAATTCA 5015

CGTTGTGTGT TATTACTTTC ACCGAATGTCT GTATTATTAA CTAAAGTATA TATTGGCAAC 5075

TAAGAAGCAA AGTGATATAA ACATGATGAC AAATTAGGCC AGGCATGGTG GCTTACTCCT 5135

30 ATAATCCCAA CATTTTGGGG GCCCAAGGTAA GGAGATCAC TTGAGGTAG GATTCAAGA 5195

CCAGCCTGAC CAACATGGTG AAACCTTGTCT TCTACTAAAA ATACAAAAT TAGCTGGGCA 5255

TGGTAGCAGG CACTCTAGT ACCAGCTACT CAGGGCTGAG GCAGGAGAAT CGCTTGAACC 5315

CAGGAGATGG AGGTGCACT GAGCTGAGAT TGTACCACTG CACTCCAGTC TGGGCAACAG 5375

AGCAAGATT CATCACACAC ACACACACAC ACACACACAC ACACATTAGA AATGTGTACT 5435

40 TGGCTTTGTT ACCTATCGTA TTAGTGCATC TATTGCATG AACTCCAAG CTACTCTCCT 5495

TGTGTTAACG TCTTCATTGG GTACAGGTCA CTACTATTAA GTTCAGGTTA TTGGATGCA 5555

TTCCACGGTA GTGATGACAA TTCATCAGGC TACTGTGTGT GTTCACCTTG TCACTCCAC 5615

CACTAGACTA ATCTCAGACC TTCACTCAA GACACATTAC ACTAAAGATG ATTTGCTTT 5675

TTGTGTTAA TCAAGCAATG GTATAAACCA GCTTGACTCT CCCCAAACAG TTTTCGTAC 5735

50 TACAAAGAAC TTATGAAGC AGAGAAATGT GAATTGATAT ATATATGAGA TTCTAACCCA 5795

GTTCCAGCAT TGTTCATG TGTAAATTGAA ATCATAGACA AGCCATTAA GCCTTGCTT 5855

55

5 TCTTATCTAA AAAAAAAA AAAAAATGA AGGAAGGGT ATTAAAAGGA GTGATCAAAT 5915
 TTTAACATTTC TCTTTAATTA ATTCACTTTT AATTTTACTT TTTTTCAATT ATTGTGCACT 5975
 TACTATCTGG TACTGTGCTA TAGAGGCTTT AACATTATA AAAACACTGT GAAAGTTGCT 6035
 10 TCAGATGAAT ATAGGTAGTA GAACGGCAGA ACTAGTATTC AAAGCCAGGT CTGATGAATC 6095
 CAAAAACAAA CACCCATTAC TCCCATTTC TGGGACATAC TTACTCTACC CAGATGCTCT 6155
 GGGCTTTGTA ATGCCTATGT AAATAACATA GTTTATGTT TGGTTATTTT CCTATGTAAT 6215
 15 GTCTACTTAT ATATCTGTAT CTATCTCTTC CTTTGTTC CAAAGTAAAC TATGTGTCTA 6275
 AATGTGGGCA AAAAATAACA CACTATTCCA AATTACTGTT CAAATTCCCT TAAGTCAGTG 6335
 ATAATTATTT GTTTGACAT TAATCATGAA GTTCCCTGTC CCTACTAGGT AAACCTTAA 6395
 20 TAGAATGTTA ATGTTTGTAT TCATTATAAG AATTGTTGGC TGTACTTAT TTACAACAAT 6455
 ATTCACTCT AATTAGACAT TTACTAAACT TTCTCTTGAA AACAAATGCCC AAAAAGAAC 6515
 ATTAGAAGAC ACGTAAGCTC AGTTGGTCTC TGCCACTAAG ACCAGCCAAC AGAACGTTGA 6575
 25 TTTTATTCAA ACTTTGCATT TTAGCATATT TTATCTTGGA AAATTCAATT GTGTTGGTTT 6635
 TTGTTTTTG TTGTATTGA ATAGACTCTC AGAAATCCAA TTGTTGACTA AATCTTCTGG 6695
 30 GTTTCTAAC CTTCTTTAG AT GTT ACC CTG TGT GAG GAG GCA TTC TTC AGG 6747

Asp Val Thr Leu Cys Glu Glu Ala Phe Phe Arg

180 185

35
 40 TTT GCT CTT CCT ACA AAG TTT ACG CCT AAC TGG CTT AGT GTC TTG GTA 6795
 Phe Ala Val Pro Thr Lys Phe Thr Pro Asn Trp Leu Ser Val Leu Val
 190 195 200

45 GAC AAT TTG CCT GGC ACC AAA GTA AAC GCA GAG AGT GTA GAG AGG ATA 6843
 Asp Asn Leu Pro Gly Thr Lys Val Asn Ala Glu Ser Val Glu Arg Ile
 50 205 210 215

AAA CGG CAA CAC AGC TCA CAA GAA CAG ACT TTC CAG CTG CTG AAG TTA 6891
 5 Lys Arg Gln His Ser Ser Gln Glu Gln Thr Phe Gln Leu Leu Lys Leu
 220 225 230 235

 10 TGG AAA CAT CAA AAC AAA GAC CAA GAT ATA GTC AAG AAG ATC ATC CAA G 6940
 15 Trp Lys His Gln Asn Lys Asp Gln Asp Ile Val Lys Ile Ile Gln
 240 245 250

 20 GTATGATAAT CTAAAATAAA AAGATCAATC AGAAATCAAA GACACCTATT TATCATAAAC 7000
 25 CAGGAACAAG ACTGCATGTA TGTAGTTG TGTGGATCTT GTTCCCTGT TGGAAATCATT 7060
 GTTGGACTGA AAAAGTTCC ACCTGATAAT GTAGATGTGA TTCCACAAAC AGTTATACAA 7120
 30 GGTTTGTTC TCACCCCTGC TCCCCAGTT CCTTGAAAG TATGTTAAC ACTCTAACAG 7180
 AAGAGAAATG CATTGAAGC CAGGGCTGTA TCTCAGGGAG TCGCTTCCAG ATCCCTAAC 7240
 GCTTCTGTAA GCAGCCCCTC TAGACCACCA AGGAGAACCT CTATAACCAC TTTGTATCTT 7300
 35 ACATTGCACC TCTACCAAGA AGCTCTGTTG TATTTACTTC GTAATTCTCT CCAGGTAGGC 7360
 TTTTCTGTAGC TTACAAATAT GTTCTTATTA ATCCTCATGA TATGCCCTGC ATTAAAATTA 7420
 TTTTAATGGC ATATGTTATG AGAATTAATG AGATAAAATC TGAAAAGTGT TTGAGCCCTCT 7480
 40 TGTAGGAAAA AGCTAGTTAC AGCAAAATGT TCTCACATCT TATAAGTTA TATAAAGATT 7540
 CTCCTTACA AATGGTGTCA GAGAGAAACA GAGAGAGATA GGGAGAGAAC TGTGAAAGAA 7600
 45 TCTGAAGAAA AGGAGTTCA TCCAGTGTGG ACTGTAAGCT TTACGACACA TGATGGAAAG 7660
 ACTTCTGACT TCAGTAAGCA TTGGGAGGAC ATGCTAGAAC AAAAGGAAG AAGAGTTCC 7720
 ATAATGCACA CAGGGTCAGT GAGAAATTCA TTCAGGTCTT CACCAAGTAGT TAAATGACTG 7780
 50 TATAGTCTTG CACTACCCCA AAAAACTTCA AGTATCTGAA ACCGGGGCAA CAGATTTAG 7840
 GAGACCAACG TCTTGAGAG CTGATTGCTT TTGCTTATCC AAAGACTAAA CTTTATGTT 7900
 TTGAGCAAAC CAAAAGTATT CTTGAAACGT ATAATTAGCC CTGAAGCCGA AAGAAAAGAG 7960
 AAAATCAGAG ACCGTTAGAA TTGGAAGCAA CCAAATTCCC TATTTATAA ATGAGGACAT 8020

TTTAACCCAG AAAGATGAAC CGATTTGGCT TAGGGCTCAC AGATACTAAG TGACTCATGT 8080
 5 CATTAAATAGA AATGTTAGTT CCTCCCTCTT AGGTTTGAC CCTAGCTTAT TACTGAAATA 8140
 TTCTCTAGGC TGTGTGTCTC CTTAGTTCC TCGACCTCAT GTCTTGAGT TTTCAGATAT 8200
 CCTCCTCATG GAGGTAGTCC TCTGGTGCTA TGTGTATTCT TAAAGGCTA GTTACGGCAA 8260
 10 TTAACCTTATC AACTAGCCGCC TACTAATGAA ACTTTGTATT ACAAAAGTAGC TAACTTGAAT 8320
 ACTTTCCCTT TTTTCTGAAA TGTATGGTG GAACTTTCTC AAACCTTTTC TTAGAAAAC 8380
 15 GAGAGTCATG TGTCTTATT TCTACTGTTA ATTTCAAAA TTAGGAGCTT CTTCCAAACT 8440
 TTTGTTGGAT GCCAAAATA TATAGCATAT TATCTTATTA TAACAAAAAA TATTTATCTC 8500
 AGTTCTTAGA AATAAAATGGT GTCACTAAC TCCCTCTCAA AACAAAAGGT TATCATTGAA 8560
 20 ATATAATTAT CAAATTCTGC AAGAACCTT TGCCTCACCG TTGTTTTATG ATGCCATTGG 8620
 ATGAATATAA ATGATGTGAA CACTTATCTG GGCTTTGCT TTATGCAG AT ATT GAC 8676

25 Asp Ile Asp

CTC TGT GAA AAC ACC GTG CAG CGG CAC ATT GGA CAT GCT AAC CTC ACC 8724
 30 Leu Cys Glu Asn Ser Val Gln Arg His Ile Gly His Ala Asn Leu Thr
 255 260 265 270

35 TTC GAG CAG CTT CGT AGC TTG ATG GAA AGC TTA CCG GGA AAG AAA GTG 8772
 Phe Glu Gln Leu Arg Ser Leu Met Glu Ser Leu Pro Gly Lys Lys Val
 40 275 280 285

45 GGA GCA GAA GAC ATT GAA AAA ACA ATA AAG GCA TGC AAA CCC AGT GAC 8820
 Gly Ala Glu Asp Ile Glu Lys Thr Ile Lys Ala Cys Lys Pro Ser Asp
 290 295 300

50 CAG ATC CTG AAG CTG CTC AGT TTG TGG CGA ATA AAA AAT GGC GAC CAA 8868
 55

Gln Ile Leu Lys Leu Leu Ser Leu Trp Arg Ile Lys Asn Gly Asp Glu

5 305 310 315

GAC ACC TTG AAC GGC CTA ATG CAC GCA CTA AAG CAC TCA AAG ACG TAC 8916

10 Asp Thr Leu Lys Gly Leu Met His Ala Leu Lys His Ser Lys Thr Tyr

320 325 330

15

CAC TTT CCC AAA ACT GTC ACT CAG AGT CTA AAG AAG ACC ATC AGG TTC 8964

His Phe Pro Lys Thr Val Thr Gln Ser Leu Lys Lys Thr Ile Arg Phe

20 335 340 345 350

25 CTT CAC AGC TTC ACA ATG TAC AAA TTG TAT CAG AAG TTA TTT TTA GAA 9012

Leu His Ser Phe Thr Met Tyr Lys Leu Tyr Gln Lys Leu Phe Leu Glu

355 360 365

30

ATG ATA GGT AAC CAG GTC CAA TCA GTA AAA ATA AGC TGC TTA 9054

35 Met Ile Gly Asn Gln Val Gln Ser Val Lys Ile Ser Cys Leu

370 375 380

40 TAACTGGAAA TGGCCATTGA GCTGTTCTCT CACAATTGGC GAGATCCCAT GGATGACTAA 9114

ACTGTTCTC AGGCACCTGA GGCTTCACT GATATCTTC TCATTACCAAG TGACTAATT 9174

45 TGCCACAGGG TACTAAAAGA AACTATGATG TGGAGAAAGG ACTAACATCT CCTCCAATAA 9234

ACCCCAAATG GTTAATCCAA CTGTCAGATC TCGATCGTTA TCTACTGACT ATATTTCCC 9294

TTATTACTGC TTGCAGTAAT TCAACTGGAA ATTAAAAAAA AAAAAGTAGA CTCCACTGGG 9354

50 CCTTACTAAA TATGGGAATG TCTAACTTAA ATAGCTTGG GATTCCAGCT ATGCTAGAGG 9414

CTTTTATTAG AAAGCCATAT TTTTTCTGT AAAAGTTACT AATATATCTG TAACACTATT 9474

55

ACAGTATTGC TATTTATATT CATTAGATA TAAGATTGG ACATATTATC ATCCTATAAA 9534
 5 GAAACGGTAT GACTAATTT TAGAAACAAA ATTATATTCT GTTATTATG ACAAATGAAA 9594
 GAGAAAATAT ATATTTTAA TGGAAACTTT GTACCATTTT TCTAATAGGT ACTGCCATAT 9654
 10 TTTCTGTGT GGAGTATTCT TATAATTAA TCTGTATAAG CTGTAATATC ATTTTATAGA 9714
 AAATGCATTA TTTAGTCAAT TGTAAATGT TCGAAAACAT ATGAAATATA AATTATCTGA 9774
 ATATTAGATG CTCTGAGAAA TTGAATGTAC CTTATTTAA AGATTTATG GTTTTATAAC 9834
 TATATAAAATG ACATTATTAAGT TTTTCAAA TTATTTTAA TTGCTTTCTC TGTGCTTTT 9894
 15 ATTT 9898

Sequence number: 3

20 Length of sequence: 401

Sequence Type: amino acid

Strandedness: single stranded

25 Topology: linear

Molecular type: protein

30

Sequence:

Met Asn Asn Leu Leu Cys Cys Ala Leu Val Phe Leu Asp Ile Ser

35 -20 -15 -10

Ile Lys Trp Thr Thr Gln Glu Thr Phe Pro Pro Lys Tyr Leu His

40 -5 1 5

Tyr Asp Glu Glu Thr Ser His Gln Leu Leu Cys Asp Lys Cys Pro

10 15 20

45 Pro Gly Thr Tyr Leu Lys Gln His Cys Thr Ala Lys Trp Lys Thr

25 30 35

50 Val Cys Ala Pro Cys Pro Asp His Tyr Tyr Thr Asp Ser Trp His

40 45 50

55

Ile Gln Asp Ile Asp Leu Cys Glu Asn Ser Val Gln Arg His Ile

5 250 255 260

Gly His Ala Asn Leu Thr Phe Glu Gln Leu Arg Ser Leu Met Glu

10 265 270 275

Ser Leu Pro Gly Lys Lys Val Gly Ala Glu Asp Ile Glu Lys Thr

15 280 285 290

Ile Lys Ala Cys Lys Pro Ser Asp Gln Ile Leu Lys Leu Leu Ser

295 300 305

20 Leu Trp Arg Ile Lys Asn Gly Asp Gln Asp Thr Leu Lys Gly Leu

310 315 320

25 Met His Ala Leu Lys His Ser Lys Thr Tyr His Phe Pro Lys Thr

325 330 335

30 Val Thr Gln Ser Leu Lys Lys Thr Ile Arg Phe Leu His Ser Phe

340 345 350

35 Thr Met Tyr Lys Leu Tyr Gln Lys Leu Phe Leu Glu Met Ile Gly

355 360 365

Asn Gln Val Gln Ser Val Lys Ile Ser Cys Leu

40 370 375 380

Sequence number: 4

45 Length of sequence: 1206

Sequence Type: nucleic acid

Strandedness: single stranded

50 Topology: linear

Molecular type: cDNA

Sequence:

5 ATGAACAACT TGCTGTGCTG CGCGCTCGTC TTTCTGGACA TCTCCATTAA GTGGACCACC 60
 CAGGAAACGT TTCCCTCCAAA CTACCTTCAT TATGACGAAG AAACCTCTCA TCAGCTGTTG 120
 10 TGTGACAAAT GTCCTCCCTGG TACCTACCTA AAACAACACT GTACAGCAAA GTGGAAGACC 180
 GTGTGCCGCC CTTGCCCTGA CCACTACTAC ACAGACAGCT GGCACACCAAG TGACCGACTGT 240
 CTATACTGCA GCCCCGTGTG CAAGGAGCTG CAGTACGTCA AGCAGGAGTC CAATCGCAC 300
 15 CACAACCCGG TGTGCGAATG CAAGGAAGGG CGCTACCTTG AGATAGAGTT CTGCTTGAAA 360
 CATAGGAGCT CCCCTCCCTGG ATTTCGAGTG GTGCAAGCTG GAACCCCAGA GCGAAATACA 420
 20 GTTGCAGAAA GATGTCCAGA TGGGTTCTTC TCAAATGAGA CCTCATCTAA AGCACCCGT 480
 AGAAAACACA CAAATTGCAG TGTCTTGGT CTCCCTGCTAA CTCAGAAAGG AAATGCAACA 540
 CACGACAACA TATGTTCCGG AAACAGTGAA TCAACTCAA AATGTGGAAT AGATGTTACC 600
 25 CTGTGTGAGG AGGCATTCTT CAGGTTTGCT GTTCCTACAA AGTTTACGCC TAACTGGCTT 660
 AGTGTCTTGG TAGACAATTG GCCTGGCACC AAAGTAAACG CAGAGAGTGT AGAGAGGATA 720
 AAACGGCAAC ACAGCTCACA AGAACAGACT TTCCAGCTGC TGAAGTTATG GAAACATCAA 780
 30 AACAAAGACC AAGATATACT CAAGAAGATC ATCCAAGATA TTGACCTCTG TGAAAACAGC 840
 GTCCAGGGGC ACATTGGACA TGCTAACCTC ACCTTCGAGC AGCTTCGTAG CTTGATGGAA 900
 35 AGCTTACCGG GAAAGAAAGT GGGAGCAGAA GACATTGAAA AAACAATAAA GGCAATGCAA 960
 CCCAGTGACC AGATCCTGAA GCTGCTCAGT TTGTGGCGAA TAAAAAATGG CGACCAAGAC 1020
 40 ACCTTGAAGG GCCTAATGCA CGCACTAAAG CACTCAAAGA CGTACCAACTT TCCCAAAACT 1080
 GTCACTCAGA GTCTAAAGAA GACCATCAGG TTCCCTTCACA GCTTCACAAT GTACAAATTG 1140
 TATCAGAACT TATTTTAAAGA AATGATAGGT AACCAAGGTCC AATCAGTAAA AATAAGCTGC 1200
 45 TTATAA 1206

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5 (i) APPLICANT:

(A) NAME: SNOW BRAND MILK PRODUCTS CO., LTD.
 (B) STREET: 1-1, NAEBOCHO 6-CHOME
 (C) CITY: HIGASHI-KU, SAPPORO-SHI
 (D) STATE: HOKKAIDO
 (E) COUNTRY: JP
 (F) POSTAL CODE (ZIP): NONE

10 (ii) TITLE OF INVENTION: NOVEL DNA AND PROCESS FOR PREPARING PROTEIN USING THE DNA

15 (iii) NUMBER OF SEQUENCES: 4

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

20 (v) CURRENT APPLICATION DATA:

APPLICATION NUMBER: EP 97935810.8

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: JP 235928/96
 (B) FILING DATE: 19-AUG-1996

25 (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1316 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA (human OCIF genomic DNA-1)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

35 CTGGAGACAT ATAACTTGAA CACTTGGCCC TGATGGGAA GCAGCTCTGC AGGGACTTTT 60
 TCAGCCATCT GTAAACAATT TCAGTGGAA CCCGCGAACT GTAATCCATG AATGGGACCA 120
 CACTTACAA GTCATCAAGT CTAACCTCTA GACCAGGGAA TTAATGGGGG AGACAGCGAA 180
 CCCTAGAGCA AAGTGCAAA CTTCTGTCGA TAGCTTGAGG CTAGTGGAAA GACCTCGAGG 240
 AGGCTACTCC AGAAGTTCAG CGCGTAGGAA GCTCCGATAC CAATAGCCCT TTGATGATGG 300
 TGGGGTTGGT GAAGGGAACA GTGCTCCGCA AGTTTATCCC TGCCCCAGGC AGTCCAATTT 360
 TCACTCTGCA GATTCTCTCT GGCTCTAACT ACCCCAGATA ACAAGGAGTG AATGCAGAAT 420
 AGCACGGGCT TTAGGGCCAA TCAGACATTA GTTAGAAAAA TTCCTACTAC ATGGTTTATG 480
 40 TAAACTTGAA GATGAATGAT TCGGAACCTCC CGCGAAAGGG CTCAGACAAAT GCCATGCATA 540
 AAGAGGGGCC CTGTAATTG AGGTTTCAGA ACCCGAAGTG AAGGGGTCAAG GCAGCCGGGT 600
 ACGGCGGAAA CTCACAGCTT TCGCCCCAGCG AGAGGACAAA GGTCTGGGAC ACACCTAAC 660
 TGCGTCCGGA TCTTGGCTGG ATCGGACTCT CAGGGTGGAG GAGACACAAG CACAGCAGCT 720
 GCCCAGCGTG TGCCCCAGCCC TCCCACCGCT GGTCCCGGCT GCCAGGAGGC TGGCCGCTGG 780
 45 CGGGAAAGGG CGGGAAACC TCAGAGCCCC GCGGAGACAG CAGCCGCCTT GTTCCTCAGC 840
 CCGGTGGCTT TTTTTTCCCC TGCTCTCCCA GGGGACAGAC ACCACCGCCC CACCCCTCAC 900
 GCCCCACCTC CCTGGGGGAT CCTTCCGCC CCAGCCCTGA AAGCGTTAAT CCTGGAGCTT 960
 TCTGCACACC CCCCGACCGC TCCCACCCAA GCTTCCCTAAA AAAGAAAGGT GCAAAGTTG 1020
 GTCCAGGATA GAAAAATGAC TGATCAAAGG CAGCGATAC TTCCTGTTGC CGGGACGCTA 1080
 TATATAACGT GATGAGCGCA CGGGCTGCGG AGACGCACCG GAGCGCTCGC CCAGCCGCG 1140
 CCTCCAAGCC CCTGAGGTTT CGGGGACCA CA ATG AAC AAG TTG CTG TGC TGC 1193
 Met Asn Lys Leu Leu Cys Cys

-20

-15

GCG CTC GTG GTAAGTCCCT GGGCCAGCCG ACGGGTGCCTGGG

1242

Ala Leu Val

GAGGCTGCTG CCACCTGGTC TCCCAACCTC CCAGCGGACC GGCGGGGAGA AGGCTCCACT 1302
 CGCTCCCTCC CAGG 1316

5

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9898 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA (human OCIF genomic DNA-2)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

15

GCTTACTTTG TGCCAAATCT CATTAGGCTT AAGGTAATAC AGGACTTTGA GTCAAATGAT 60
 ACTGTTGCAC ATAAGAACAA ACCTATTTTC ATGCTAAGAT GATGCCACTG TGTTCCCTTC 120
 TCCTTCTAG TTT CTG GAC ATC TCC ATT AAG TGG ACC ACC CAG GAA ACG TTT 171
 Phe Leu Asp Ile Ser Ile Lys Trp Thr Thr Gln Glu Thr Phe
 -10 -5 1

20

CCT CCA AAG TAC CTT CAT TAT GAC GAA GAA ACC TCT CAT CAG CTG TTG 219
 Pro Pro Lys Tyr Leu His Tyr Asp Glu Glu Thr Ser His Gln Leu Leu
 5 10 15

25

TGT GAC AAA TGT CCT CCT GGT ACC TAC CTA AAA CAA CAC TGT ACA GCA 267
 Cys Asp Lys Cys Pro Pro Gly Thr Tyr Leu Lys Gln His Cys Thr Ala
 20 25 30 35

30

AAG TGG AAG ACC GTG TGC GCC CCT TGC CCT GAC CAC TAC TAC ACA GAC 315
 Lys Trp Lys Thr Val Cys Ala Pro Cys Pro Asp His Tyr Tyr Thr Asp
 40 45 50

35

AGC TGG CAC ACC AGT GAC GAG TGT CTA TAC TGC AGC CCC GTG TGC AAG 363
 Ser Trp His Thr Ser Asp Glu Cys Leu Tyr Cys Ser Pro Val Cys Lys
 55 60 65

40

GAG CTG CAG TAC GTC AAG CAG GAG TGC AAT CGC ACC CAC AAC CGC GTG 411
 Glu Leu Gln Tyr Val Lys Gln Glu Cys Asn Arg Thr His Asn Arg Val
 70 75 80

45

TGC GAA TGC AAG GAA GGG CGC TAC CTT GAG ATA GAG TTC TGC TTG AAA 459
 Cys Glu Cys Lys Glu Gly Arg Tyr Leu Glu Ile Glu Phe Cys Leu Lys
 85 90 95

50

CAT AGG AGC TGC CCT CCT GGA TTT GGA GTG GTG CAA GCT G GTACGTGTCA 509
 His Arg Ser Cys Pro Pro Gly Phe Gly Val Val Gln Ala
 100 105 110

55

ATGTGCAGCA AAATTAATTA GGATCATGCA AAGTCAGATA GTTGTGACAG TTTAGGAGAA 569
 CACTTTGTT CTGATGACAT TATAGGATAG CAAATTGCAA AGGTAAATGAA ACCTGCCAGG 629
 TAGGTACTAT GTGTCTGGAG TGCTTCCAAA GGACCATTGC TCAGAGGAAT ACTTTGCCAC 689
 TACAGGGCAA TTTAATGACA AATCTCAAAT GCAGCAAATT ATTCTCTCAT GAGATGCATG 749
 ATGGTTTTTT TTTTTTTTTT TAAAGAAAACA AACTCAAGTT GCACTATTGA TAGTTGATCT 809
 ATACCTCTAT ATTCACCTTC AGCATGGACA CCTTCAAACT GCAGCATTG TTGACAAACA 869
 TCAGAAAATGT TAATTATAC CAAGAGAGTA ATTATGCTCA TATTATGAG ACTCTGGAGT 929
 GCTAACAAATA AGCAGTTATA ATTAATTATG TAAAAAATGA GAATGGTGAG GGGAAATTGCA 989
 TTTCATTATT AAAAACAAAGG CTAGTTCTTC CTTTAGCATG GGAGCTGAGT GTTTGGGAGG 1049
 GTAAGGACTA TAGCAGAACATC TCTTCATGCA GCTTATTCTT TATCTTAGAC AAAACAGATT 1109
 GTCAAGCCAA GAGCAAGCAC TTGCTTATAA ACCAAGTGCT TTCTCTTTG CATTGAAAC 1169
 AGCATTGGTC AGGGCTCATG TGTATTGAAT CTTTTAAACC AGTAACCCAC GTTTTTTTTC 1229
 TGCCACATTT GCGAACGTTTC AGTGCAGCCT ATAACTTTTC ATAGCTTGAG AAAATTAAGA 1289
 GTATCCACTT ACTTAGATGG AAGAAGTAAT CAGTATAGAT TCTGATGACT CAGTTGAG 1349

55

5 CAGTGTCT CAACTGAAGC CCTGCTGATA TTTTAAGAAA TATCTGGATT CCTAGGCTGG 1409
 ACTCCTTTT GTGGGCAGCT GTCCTGCAC TTGAGAATT TTGGCAGCAC CCCTGGACTC 1469
 TAGCCACTAG ATACCAATAG CAGTCCTCC CCCATGTGAC AGCCAAAAT GTCTTCAGAC 1529
 ACTGTCAAAT GTGCCAGGT GGCAAAATCA CTCTGGTTG AGAACAGGGT CATCAATGCT 1589
 AAGTATCTGT AACTATTTA ACTCTCAAAA CTTGTGATAT ACAAAGTCTA AATTATTAGA 1649
 CGACCAATAC TTTAGGTTA AAGGCATACA AATGAAACAT TCAAAATCA AAATCTATT 1709
 TGTTTCTCAA ATAGTGAATC TTATAAAATT AATCACAGAA GATGCAAATT GCATCAGAGT 1769
 CCCTTAAAT TCTCTTCGT ATGAGTATTT GAGGGAGGAA TTGGTGATAG TTCTACTTT 1829
 CTATTGGATG GTACTTTGAG ACTCAAAAGC TAAGCTAAGT TGTGTGTG TGCG 1889
 GGGGTGGAA TCCCCTCAGA TAAAAGCAA TCCATGTAAT TCATTCAAGT AGTTGTATAT 1949
 10 GTAGAAAAAT GAAAAGTGGG CTATGCAGCT TGGAAACTAG AGAATTTGAA AAAATAATGG 2009
 AAATCACAAG GATCTTTCTT AAATAAGTAA GAAAATCTGT TTGAGAATG AAGCAAGCAG 2069
 GCAGCCAGAA GACTCAGAAC AAAAGTACAC ATTACTCT GTGTACACTG GCAGCACAGT 2129
 GGGATTATTACCTCTCCC TCCCTAAAAA CCCACACAGC GGTTCTCTT GGGAAATAAG 2189
 AGGTTCCAG CCCAAAGAGA AGGAAAGACT ATGTTGGTGT ACTCTAAAAA GTATTTAATA 2249
 15 ACCGTTTTGT TGTTGCTGTT GCTGTTGA AATCAGATTG TCTCCTCTCC ATATTTATT 2309
 TACTTCATTC TGTTAATTCC TGTGGAATTA CTTAGAGCAA GCATGGTGAA TTCTCAACTG 2369
 TAAAGCCAAA TTTCTCCATC ATTATAATT CACATTTGTC CTGGCAGGTT ATAATTTTA 2429
 TATTTCACAT GATAGTAATA AGGTAATACT ATTACTTAGA TGGATAGATC TTTTCATAA 2489
 AAAGTACCAT CAGTTATAGA GGGAAAGTCAT GTTCACTGTT AGGAAGGTCA TTAGATAAAG 2549
 CTTCTGAATA TATTATGAAA CATTAGTTCT GTCAATTCTTA GATTCTTTT GTAAATAAAC 2609
 TTTAAAGCT AACTTACCTA AAAGAAATAT CTGACACATA TGAACCTCT ATTAGGATGC 2669
 20 AGGAGAACAG CCAAGCCACA GATATGTATC TGAAGAATGA ACAAGATTCT TAGGCCCGGC 2729
 ACGGTGGCTC ACATCTGTAA TCTCAAGAGT TTGAGAGGTC AAGGCGGGCA GATCACCTGA 2789
 GGTCAGGAGT TCAAGACCAG CCTGGCAAC ATGATGAAAC CCTGCCTCTA CTAAAATAC 2849
 AAAAATTAGC AGGGCATGGT GGTGCATGCC TGCAACCCCTA GCTACTCAGG AGGCTGAGAC 2909
 AGGAGAATCT CTTGAACCCCT CGAGGCCGGAG GTTGTGGTGA GCTGAGATCC CTCTACTGCA 2969
 CTCCAGCCTG GGTGACAGAG ATGAGACTCC GTCCCTGCCG CCGCCCCCGC CTTCCCCCCC 3029
 AAAAGATTCT TTCTTCATGC AGAACATACG GCAGTCACAA AAGGGAGACC TGGGTCCAGG 3089
 25 TGTCCAAGTC ACTTATTTG AGTAAATTAG CAATGAAAGA ATGCCATGGA ATCCCTGCC 3149
 AAATAACCTCT GCTTATGATA TTGAGAATT TGATATAGAG TTGTATCCCA TTTAAGGAGT 3209
 AGGATGTAGT AGGAAAGTAC TAAAACAAA CACACAAAACA GAAAACCTC TTTGCTTTGT 3269
 AAGGTGGTTC CTAAGATAAT GTCACTGCAA TGCTGGAAAT AATATTTAAT ATGTGAAGGT 3329
 TTTAGGCTGT GTTTTCCCCCT CCTGTTCTTT TTTCTGCCA GCCCTTGTC ATTTTTGCAG 3389
 GTCAATGAAT CATGTAGAAA GAGACAGGAG ATGAAACTAG AACCACTCCA TTTTGCCCT 3449
 30 TTTTTTATTT TCTGGTTTG GTAAAAGATA CAATGAGGTA GGAGGTTGAG ATTTATAAAT 3509
 GAAGTTTAAT AAGTTTCTGT AGCTTGATT TTCTCTTTT ATATTTGTTA TCTTCATCAA 3569
 GCCAGAATTG GCCTGTAAAAA TCTACATATG GATATTGAAG TCTAAATCTG TTCAACTAGC 3629
 TTACACTAGA TGGAGATATT TTCAATTCA GATACACTGG AATGTATGAT CTAGCCATGC 3689
 GTAATATAGT CAAGTGTGTT AAGGTATTAA TTTTTAATAG CGTCCTTAGT TGTGGACTGG 3749
 TTCAAGTTTT TCTGCCATG ATTTCTTCAA ATTTATCAA TATTTTCCCA TCATGAAGTA 3809
 35 AAATGCCCTT GCAGTCACCC TTCTGAAGT TTGAAACGACT CTGCTGTTT AAACAGTTA 3869
 AGCAAATGGT ATATCATCTT CCGTTACTA TGAGCTTAA CTGCAGGCTT ACGCTTTGA 3929
 GTCAGCGGCC AACTTTATTG CCACCTCAA AAGTTTATTA TAATGTTGTA AATTTTACT 3989
 TCTCAAGGTT AGCATACTTA GGAGTTGCTT CACAATTAGG ATTCAAGGAA GAAAGAAACTT 4049
 CAGTAGGAAC TGATTGGAAT TTAATGATGC AGCATTCAAT GGGTACTAAT TTCAAAGAAT 4109
 GATATTACAG CAGACACACA GCAGTTATCT TGATTTCTA GGAATAATTG TATGAAGAAT 4169
 40 ATGGCTGACA ACACGGCCTT ACTGCACATC AGCGGGAGGCT GGACTAATGA ACACCCCTACC 4229
 CTTCTTCTCCT TTCTCTCAC ATTTCATGAG CGTTTTGTAG GTAAACGAGAA AATTGACTTG 4289
 CATTGCAATT ACAAGGAGGA GAAACTGGCA AAGGGGATGA TGGTGGAAAGT TTTGTTCTGT 4349
 CTAATGAAGT GAAAATGAA AATGCTAGAG TTTTGTGCAA CATAATAGTA GCAGTAAAAA 4409
 CCAAGTGAAAGTCTTCA AACTGTGTT AAGAGGGCAT CTGCTGGAA ACGATTTGAG 4469
 GAGAAGGTAC TAAATTGCTT GGTATTTCC GTAG GA ACC CCA GAG CGA AAT ACA 4523
 Gly Thr Pro Glu Arg Asn Thr
 45 115

50	GTT TGC AAA AGA TGT CCA GAT GGG TTC TTC TCA AAT GAG ACG TCA TCT	4571
	Val Cys Lys Arg Cys Pro Asp Gly Phe Phe Ser Asn Glu Thr Ser Ser	
	120 125 130 135	
50	AAA GCA CCC TGT AGA AAA CAC ACA AAT TGC AGT GTC TTT GGT CTC CTG	4619
	Lys Ala Pro Cys Arg Lys His Thr Asn Cys Ser Val Phe Gly Leu Leu	
	140 145 150	
	CTA ACT CAG AAA GGA AAT GCA ACA CAC GAC AAC ATA TGT TCC GGA AAC	4667

Leu Thr Gln Lys Gly Asn Ala Thr His Asp Asn Ile Cys Ser Gly Asn
 155 160 165

5 AGT GAA TCA ACT CAA AAA TGT GGA ATA G GTAATTACAT TCCAAAATAC 4715
 Ser Glu Ser Thr Gln Lys Cys Gly Ile
 170 175

10 GTCTTTGTAC GATTTGTAG TATCATCTCT CTCTCTGAGT TGAACACAAG GCCTCCAGCC 4775
 ACATTCTTGG TCAAACTTAC ATTTCCCTT TCTTGAATCT TAACCAAGCTA AGGCTACTCT 4835
 CGATGCATTA CTGCTAAAGC TACCACTCAG AATCTCTCAA AAACCTCATCT TCTCACAGAT 4895
 AACACCTCAA AGCTTGATTT TCTCTCCTT CACACTGAAA TCAAAATCTTG CCCATAGGCA 4955
 AAGGGCAGTG TCAAGTTG CACTGAGAT AAATTAGGAG AGTCCAAACT GTAGAATTCA 5015
 CGTTGTGTGT TATTACTTTC ACGAATGTCT GTATTATTAA CTAAAGTATA TATTGGCAAC 5075
 TAAGAAGCAA AGTGATATAA ACATGATGAC AAATTAGGCC AGGCATGGTG GCTTACTCCT 5135
 ATAATCCCAA CATTGGGGGG GGCCAGGTA GGCGAGATCAC TTGAGGTAG GATTTCAGA 5195
 CCAGCCTGAC CAACATGGT AAACCTTGTCT TCTACTAAAAT ATACAAAAT TAGCTGGCA 5255
 TGGTAGCAGG CACTTCTAGT ACCAGCTACT CAGGGCTGAG GCAGGAGAAT CGCTTGAAAC 5315
 CAGGAGATGG AGGTTGCAGT GAGCTGAGAT TGTACCACTG CACTCCAGTC TGGGCAACAG 5375
 AGCAAGATTT CATCACACAC ACACACACAC ACACACACAC ACACATTAGA AATGTGTACT 5435
 TGGCTTTGTT ACCTATGGTA TTAGTGCATC TATTGCATGG AACTTCCAAG CTACTCTGGT 5495
 TGTGTTAACG TCTTCATTGG GTACAGGTCA CTAGTATTAA GTTCAGGTTA TTGGGATGCA 5555
 TTCCACGGTA GTGATGACAA TTCATCAGGC TAGTGTGTGT GTTCACCTTG TCACTCCCAC 5615
 20 CACTAGACTA ATCTCAGACC TTCACTCAA GACACATTAC ACTAAAGATG ATTTGCTTT 5675
 TTGTGTTAA TCAAGCAATG GTATAAACCA GCTTGACTCT CCCCAAACAG TTTTCGTAC 5735
 TACAAAGAAG TTTATGAAGC AGAGAAATGT GAATTGATAT ATATATGAGA TTCTAACCCA 5795
 GTTCCAGCAT TGTTCATTG TGTAAATTGAA ATCATAGACA AGCCATTAA GCCTTGCTT 5855
 TCTTATCTAA AAAAAAAA AAAAAAATGA AGGAAGGGGT ATTAAAAGGA GTGATCAAAT 5915
 TTTAACATTC TCTTTAATTA ATTCAATTAA AATTTTACTT TTTTCATTT ATTGTGCACT 5975
 TACTATGTGG TACTGTGCTA TAGAGGCTTT AACATTTATA AAAACACTGT GAAAGTTGCT 6035
 25 TCAGATGAAT ATAGGTAGTA GAAAGGCAAG ACTAGTATTC AAAGCCAGGT CTGATGAATC 6095
 CAAAAACAAA CACCCATTAC TCCCATTTC TGGGACATAC TTACTCTTAC CAGATGCTCT 6155
 GGGCTTTGTA ATGCCTATGT AAATAACATA GTTTTATGTT TGGTATTCTT CCTATGTAAT 6215
 GTCTACTTTAT ATATCTGTAT CTATCTCTTG CTTTGTTCC AAAGGTAAAC TATGTGTCTA 6275
 AATGTGGGCA AAAATAACA CACTATTCCA AATTACTGTT CAAATTCCCTT TAAGTCAGTG 6335
 ATAATTATTGTTTTGACAT TAATCATGAA GTTCCCTGTG GGTACTAGGT AACCTTTAA 6395
 TAGAATGTTA ATGTTTGAT TCATTATAAG AATTTTGGC TGTTACTTAT TTACAACAAAT 6455
 30 ATTTCACTCT AATTAGACAT TTACTAAACT TTCTCTTGAA AACAAATGCC AAAAAAGAAC 6515
 ATTAGAAGAC ACGTAAGCTC AGTTGGTCTC TGCCACTAAG ACCAGCCAAC AGAAGCTTGA 6575
 TTTTATTCAA ACTTTGCTTAC TTAGCATATT TTATCTTGGAA AAATTCAATT GTGTTGGTTT 6635
 TTTGTTTTTG TTTGTATTGA ATAGACTCTC AGAAATCCTAA TTGTTGAGTA AATCTCTGG 6695
 GTTTCTAAC CTTCTTTAG AT GTT ACC CTG TGT GAG GAG GCA TTC TTC AGG 6747
 Asp Val Thr Leu Cys Glu Glu Ala Phe Phe Arg
 35 180 185

TTT GCT GTT CCT ACA AAG TTT ACG CCT AAC TGG CTT AGT GTC TTG GTA 6795
 Phe Ala Val Pro Thr Lys Phe Thr Pro Asn Trp Leu Ser Val Leu Val
 190 195 200

40 GAC AAT TTG CCT GGC ACC AAA GTA AAC GCA GAG AGT GTA GAG AGG ATA 6843
 Asp Asn Leu Pro Gly Thr Lys Val Asn Ala Glu Ser Val Glu Arg Ile
 205 210 215

45 AAA CGG CAA CAC AGC TCA CAA GAA CAG ACT TTC CAG CTG CTG AAG TTA 6891
 Lys Arg Gln His Ser Ser Gln Glu Gln Thr Phe Gln Leu Leu Lys Leu
 220 225 230 235

TGG AAA CAT CAA AAC AAA GAC CAA GAT ATA GTC AAG AAG ATC ATC CAA G 6940
 Trp Lys His Gln Asn Lys Asp Gln Asp Ile Val Lys Lys Ile Ile Gln
 240 245 250

50 GTATGATAAT CTAAAATAAA AAGATCAATC AGAAATCAA GACACCTATT TATCATAAAC 7000
 CAGGAACAAG ACTGCATGTA TGTGGATCTT GTGGATCTT GTGGATCTT TGGAATCATT 7060
 GTTGGACTGA AAAAGTTCC ACCTGATAAT GTAGATGTGA TTCCACAAAC AGTTATACAA 7120
 GGTTTTGTTC TCACCCCTGC TCCCCAGTTT CTTGTAAAG TATGTTGAC ACTCTAAGAG 7180
 AAGAGAAATG CATTGAAGG CAGGGCTGTA TCTCAGGGAG TCGCTTCCAG ATCCCTTAAC 7240

5 GCTTCTGTAA GCAGCCCCCTC TAGACCACCA AGGAGAAGCT CTATAACCAC TTTGTATCTT 7300
 ACATTCGACC TCTACCAAGA AGCTCTGTTG TATTTACTTG GTAATTCTCT CCAGGTAGGC 7360
 TTTTCGTAGC TTACAAATAT GTTCTTATTG ATCCCTCATGA TATGGCCTGC ATTAAAATTA 7420
 TTTTAATGGC ATATGTTATG AGAATTAATG AGATAAAATC TGAAAAGTGT TTGAGCCTCT 7480
 TGTAGGAAAA AGCTAGTTAC AGCAAAATGT TCTCACATCT TATAAGTTA TATAAAGATT 7540
 CTCCTTACA AATGGGTGAA GAGAGAAACA GAGAGAGATA GGGAGAGAAG TGTGAAAGAA 7600
 TCTGAAGAAA AGGAGTTTC TCCAGTGTGG ACTGTAAGCT TTACGACACA TGATGGAAAG 7660
 AGTTCTGACT TCAGTAAGCA TTGGGAGGAC ATGCTAGAAG AAAAGGAAG AAGAGTTCC 7720
 ATAATGCAGA CAGGGTCAGT GAGAAATTCA TTCAAGGCTCT CACCAAGTAGT TAAATGACTG 7780
 10 TATAGTCTTG CACTACCCCTA AAAAATTCA AGTATCTGAA ACCGGGGCAA CAGATTCTAG 7840
 GAGACCAACG TCTTTGAGAG CTGATTGCTT TTGCTTATGC AAAGAGTAA CTTTTATGTT 7900
 TTGAGCAAC CAAAAGTATT CTTTGAAACGT ATAATTAGCC CTGAAGCCGA AAGAAAAGAG 7960
 AAAATCAGAG ACCGTTAGAA TTGGAAGCAA CCAAATTCCC TATTTTATAA ATGAGGACAT 8020
 TTTAACCCAG AAGATGAAC CGATTTGGCT TAGGGCTCAC AGATACTAAG TGACTCATGT 8080
 CATTAATAGA AATGTTAGTT CCTCCCTCTT AGGTTTGTAC CCTAGCTTAT TACTGAAATA 8140
 TTCTCTAGGC TGTGTGTCTC CTTTAGTTCC TCGACCTCAT GTCTTTGAGT TTTCAGATAT 8200
 CCTCCTCATG GAGGTAGTCC TCTGGTGTCA TGTTGTTCTT TTAAAGGCTA GTTACGGCAA 8260
 TTAACTTATC AACTAGCGCC TACTAATGAA ACTTTGATT ACAAAAGTAGC TAACTTGAAT 8320
 ACITTCCTTT TTTTCTGAAA TGTTATGGTG GTAAATTCTC AAACTTTTTC TTAGAAAAGT 8380
 GAGAGTGATG TGTCTTATTG TCTACTGTTA ATTTCAAAAA TTAGGAGCTT CTTCCAAAGT 8440
 TTTGTTGGAT GCCAAAAATA TATAGCATAT TATCTTATTA TAACAAAAAA TATTATCTC 8500
 AGTTCTTAGA AATAAATGGT GTCACTTAAC TCCCTCTCAA AAGAAAAGGT TATCATTGAA 8560
 15 20 ATATAATTAT GAAATTCTGC AAGAACCTT TGCTCTCACGC TTGTTTATG ATGGCATTGG 8620
 ATGAATATAA ATGATGTGAA CACTTATCTG GGCTTTGCT TTATGCAG AT ATT GAC 8676
 Asp Ile Asp

25 CTC TGT GAA AAC AGC GTG CAG CGG CAC ATT GGA CAT GCT AAC CTC ACC 8724
 Leu Cys Glu Asn Ser Val Gln Arg His Ile Gly His Ala Asn Leu Thr
 255 260 265 270

TTC GAG CAG CTT CGT AGC TTG ATG GAA AGC TTA CCG GGA AAG AAA GTG 8772
 Phe Glu Gln Leu Arg Ser Leu Met Glu Ser Leu Pro Gly Lys Lys Val
 275 280 285

30 GGA GCA GAA GAC ATT GAA AAA ACA ATA AAG GCA TGC AAA CCC AGT GAC 8820
 Gly Ala Glu Asp Ile Glu Lys Thr Ile Lys Ala Cys Lys Pro Ser Asp
 290 295 300

CAG ATC CTG AAG CTG CTC AGT TTG TGG CGA ATA AAA AAT GGC GAC CAA 8868
 Gln Ile Leu Lys Leu Leu Ser Leu Trp Arg Ile Lys Asn Gly Asp Gln
 305 310 315

35 GAC ACC TTG AAG GGC CTA ATG CAC GCA CTA AAG CAC TCA AAG ACG TAC 8916
 Asp Thr Leu Lys Gly Leu Met His Ala Leu Lys His Ser Lys Thr Tyr
 320 325 330

40 CAC TTT CCC AAA ACT GTC ACT CAG AGT CTA AAG AAG ACC ATC AGG TTC 8964
 His Phe Pro Lys Thr Val Thr Gln Ser Leu Lys Lys Thr Ile Arg Phe
 335 340 345 350

CTT CAC AGC TTC ACA ATG TAC AAA TTG TAT CAG AAG TTA TTT TTA GAA 9012
 Leu His Ser Phe Thr Met Tyr Lys Leu Tyr Gln Lys Leu Phe Leu Glu
 355 360 365

45 ATG ATA GGT AAC CAG GTC CAA TCA GTA AAA ATA AGC TGC TTA 9054
 Met Ile Gly Asn Gln Val Gln Ser Val Lys Ile Ser Cys Leu
 370 375 380

50 TAACTGGAAA TGGCCATTGA GCTGTTCCCT CACAATTGGC GAGATCCCAT GGATGAGTAA 9114
 ACTGTTCTC AGGCACATTGA GGCTTTCAGT GATATCTTTC TCATTACAG TGACTAATT 9174
 TGCCACAGGG TACTAAAAGA AACTATGATG TGGAGAAAGG ACTAACATCT CCTCCAATAA 9234
 ACCCCAAATG GTTAATCCAA CTGTCAGATC TGGATCGTTA TCTACTGACT ATATTTCCC 9294
 TTATTACTGC TTGCAGTAAT TCAACTGGAA ATTAAAAAAA AAAAACTAGA CTCCACTGGG 9354
 CCTTACTAAA TATGGGAATG TCTAACTTAA ATAGCTTTGG GATTCCAGCT ATGCTAGAGG 9414
 CTTTTATTAG AAAGCCATAT TTTTTCTGT AAAAGTTACT AATATATCTG TAACACTATT 9474

5
 ACAGTATTGC TATTTATATT CATTAGATA TAAGATTGG ACATATTATC ATCCTATAAA 9534
 GAAACGGTAT GACTAATTT TAGAAAGAAA ATTATATTCT GTTTATTATG ACAATGAAA 9594
 GAGAAATAT ATATTTTAA TGAAAGTTT GTAGCATTGT TCTAATAGGT ACTGCCATAT 9654
 TTTTCTGTGT GGAGTATTT TATAATTTA TCTGTATAAG CTGTAATATC ATTTTATAGA 9714
 AAATGCATTA TTTAGTCAT TGTAAATATGT TGGAAAACAT ATGAAATATA AATTATCTGA 9774
 ATATTAGATG CTCTGAGAAA TTGAATGTAC CTTATTTAAA AGATTTATG GTTTTATAAC 9834
 TATATAAATG ACATTATCAA AGTTTCAA TTATTTTA TTGCTTCTC TGTTGCTTT 9894
 ATTT 9898

10 (2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 401 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 15 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Asn Asn Leu Leu Cys Cys Ala Leu Val Phe Leu Asp Ile Ser
 20 -20 -15 -10
 Ile Lys Trp Thr Thr Gln Glu Thr Phe Pro Pro Lys Tyr Leu His
 -5 -5 1 5
 Tyr Asp Glu Glu Thr Ser His Gln Leu Leu Cys Asp Lys Cys Pro
 10 15 20
 Pro Gly Thr Tyr Leu Lys Gln His Cys Thr Ala Lys Trp Lys Thr
 25 30 35
 25 Val Cys Ala Pro Cys Pro Asp His Tyr Tyr Asp Ser Trp His
 40 45 50
 Thr Ser Asp Glu Cys Leu Tyr Cys Ser Pro Val Cys Lys Glu Leu
 55 60 65
 Gln Tyr Val Lys Gln Glu Cys Asn Arg Thr His Asn Arg Val Cys
 70 75 80
 30 Glu Cys Lys Glu Gly Arg Tyr Leu Glu Ile Glu Phe Cys Leu Lys
 85 90 95
 His Arg Ser Cys Pro Pro Gly Phe Gly Val Val Gln Ala Gly Thr
 100 105 110
 Pro Glu Arg Asn Thr Val Cys Lys Arg Cys Pro Asp Gly Phe Phe
 115 120 125
 Ser Asn Glu Thr Ser Ser Lys Ala Pro Cys Arg Lys His Thr Asn
 130 135 140
 35 Cys Ser Val Phe Gly Leu Leu Thr Gln Lys Gly Asn Ala Thr
 145 150 155
 His Asp Asn Ile Cys Ser Gly Asn Ser Glu Ser Thr Gln Lys Cys
 160 165 170
 Gly Ile Asp Val Thr Leu Cys Glu Ala Phe Phe Arg Phe Ala
 175 180 185
 40 Val Pro Thr Lys Phe Thr Pro Asn Trp Leu Ser Val Leu Val Asp
 190 195 200
 Asn Leu Pro Gly Thr Lys Val Asn Ala Glu Ser Val Glu Arg Ile
 205 210 215
 Lys Arg Gln His Ser Ser Gln Glu Gln Thr Phe Gln Leu Leu Lys
 220 225 230
 45 Leu Trp Lys His Gln Asn Lys Asp Gln Asp Ile Val Lys Lys Ile
 235 240 245
 Ile Gln Asp Ile Asp Leu Cys Glu Asn Ser Val Gln Arg His Ile
 250 255 260
 Gly His Ala Asn Leu Thr Phe Glu Gln Leu Arg Ser Leu Met Glu
 265 270 275
 50 Ser Leu Pro Gly Lys Lys Val Gly Ala Glu Asp Ile Glu Lys Thr
 280 285 290
 Ile Lys Ala Cys Lys Pro Ser Asp Gln Ile Leu Lys Leu Leu Ser
 295 300 305

	Leu	Trp	Arg	Ile	Lys	Asn	Gly	Asp	Gln	Asp	Thr	Leu	Lys	Gly	Leu
	310					315					320				
5	Met	His	Ala	Leu	Lys	His	Ser	Lys	Thr	Tyr	His	Phe	Pro	Lys	Thr
	325					330					335				
	Val	Thr	Gln	Ser	Leu	Lys	Lys	Thr	Ile	Arg	Phe	Leu	His	Ser	Phe
	340					345					350				
	Thr	Met	Tyr	Lys	Leu	Tyr	Gln	Lys	Leu	Phe	Leu	Glu	Met	Ile	Gly
	355					360					365				
10	Asn	Gln	Val	Gln	Ser	Val	Lys	Ile	Ser	Cys	Leu				
	370					375					380				

(2) INFORMATION FOR SEO ID NO:4:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1206 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

25	ATGAAACA CAGGAAAC TGTGACA GTGTGCG CTATACTG CACAAACG CATAGGAG GTTTGCA AGAAAAC CACGACA CTGTGTG AGTGTCT AAACGGCA AACAAAGA GTGCAGCG AGCTTACCG CCCAGTGAC ACCTTGAAG GTCACTCAG TATCAGAA TTATATAA	ACT TGCTGTGCTG GT CAGGAAAC AT TGTGACA CC GTGTGCG CA CTATACTG CG CACAAACG CT CATAGGAG GG GCTCTCTG AA ATTGGAG AA GATGTCC AA CAAATTG AA TATGTTCC AA AGGCAATT AA GAGCTTC AA AACAGTGA AA CAGGTTG AA TCAACTC AA TCAACT AA GTGGAAT AA AGATAGAG AA GTGCTTGA AA GACCCCAG AA CGTCACTAA AA CTCAGAAAG AA AAATGCAACA AA AATGTTACC AA AGATGTTA AA AGTGTGGCT AA AGCACCCAG AA CGTCATCTA AA CTCAGAAAG AA AAATGCAACA AA AATGTTACC AA AGATGTTA AA AGAGAGGATA AA TGAAGTTATG AA GAAACATCAA AA TTGACCTCTG AA TGAAAACAGC AA AGCTTCGAG AA AGCTTACCG AA GACATTGAA AA AAACAAATAA AA GGGAGCAGA AA GACATTGAA AA TTGTGGCGAA AA TAAAAAATGG AA CGTACCAACTT AA TCCCAAAC AA CGTACCAACTT AA GCTTCACA AA AATCAGTAA AA AATAAGCTGC AA 1206
30	TTCTGGACA TATGACGAAG AAACAAACACT ACAGACAGCT CAGTACGTCA CGCTACCTTG AGATAGAGTT GTGCAAGCTG GGGGTTCTTC TCAAATGAGA CTCCTGCTAA TCAACTCAA AATGTGGAAT AGTGTGGCTT GAACCCCAGA GCGAAATACA CTCAGAAAGG AAATGCAACA AATGTGGAAT AGATGTTACC AGTGTGGCTT AACTGGCTT CAGAGAGTGT AGAGAGGATA TGAAGTTATG GAAACATCAA TTGACCTCTG TGAAAACAGC CAGAGAGTGT AGAGAGGATA TGAAGTTATG GAAACATCAA TTGACCTCTG TGAAAACAGC CAGAGAGTGT AGAGAGGATA TGAAGTTATG GAAACATCAA TTGACCTCTG TGAAAACAGC AGCTTACCG GAAAGAAAGT GGGAGCAGAA GACATTGAA AAACAAATAA GGCATGCAAA TAAAAAATGG CGACCAAAGAC TTGTGGCGAA TAAAAAATGG CGACCAAAGAC TCCCAAAC CGTACCAACTT TCCCAAAC GCTTCACA AATCAGTAA AATAAGCTGC 1206	
35	TTCTGGACA TATGACGAAG AAACAAACACT ACAGACAGCT CAGTACGTCA CGCTACCTTG AGATAGAGTT GTGCAAGCTG GGGGTTCTTC TCAAATGAGA CTCCTGCTAA TCAACTCAA AATGTGGAAT AGATGTTACC AGTGTGGCTT AACTGGCTT CAGAGAGTGT AGAGAGGATA TGAAGTTATG GAAACATCAA TTGACCTCTG TGAAAACAGC CAGAGAGTGT AGAGAGGATA TGAAGTTATG GAAACATCAA TTGACCTCTG TGAAAACAGC AGCTTACCG GAAAGAAAGT GGGAGCAGAA GACATTGAA AAACAAATAA GGCATGCAAA TAAAAAATGG CGACCAAAGAC TTGTGGCGAA TAAAAAATGG CGACCAAAGAC TCCCAAAC CGTACCAACTT TCCCAAAC GCTTCACA AATCAGTAA AATAAGCTGC 1206	
40	TTCTGGACA TATGACGAAG AAACAAACACT ACAGACAGCT CAGTACGTCA CGCTACCTTG AGATAGAGTT GTGCAAGCTG GGGGTTCTTC TCAAATGAGA CTCCTGCTAA TCAACTCAA AATGTGGAAT AGATGTTACC AGTGTGGCTT AACTGGCTT CAGAGAGTGT AGAGAGGATA TGAAGTTATG GAAACATCAA TTGACCTCTG TGAAAACAGC CAGAGAGTGT AGAGAGGATA TGAAGTTATG GAAACATCAA TTGACCTCTG TGAAAACAGC AGCTTACCG GAAAGAAAGT GGGAGCAGAA GACATTGAA AAACAAATAA GGCATGCAAA TAAAAAATGG CGACCAAAGAC TTGTGGCGAA TAAAAAATGG CGACCAAAGAC TCCCAAAC CGTACCAACTT TCCCAAAC GCTTCACA AATCAGTAA AATAAGCTGC 1206	

Claims

50 1. A DNA comprising the nucleotide sequences of the Sequences No. 1 and No. 2 in the Sequence Table.

2. The DNA according to claim 1, wherein the Sequence ID No. 1 includes the first exon of the OCIF gene and the Sequence ID No. 2 includes the second, third, fourth, and fifth exons.

55 3. A protein exhibiting the activity of inhibiting differentiation and/or maturation of osteoclasts and having the following physicochemical characteristics,

(a) molecular weight (SDS-PAGE):

- (i) Under reducing conditions: about 60 kD,
- (ii) Under non-reducing conditions: about 60 kD and about 120 kD;

5 (b) amino acid sequence:
includes an amino acid sequence of the Sequence ID No. 3 in the Sequence Table,

(c) affinity:
exhibits affinity to a cation exchanger and heparin, and

(d) heat stability:

10 (i) the osteoclastogenesis-inhibitory activity is reduced when treated with heat at 70°C for 10 minutes or at 56°C for 30 minutes,
(ii) the osteoclastogenesis-inhibitory activity is lost when treated with heat at 90°C for 10 minutes.

15 4. A process for producing a protein exhibiting an activity of inhibiting differentiation and/or maturation of osteoclasts and having the following physicochemical characteristics,

(a) molecular weight (SDS-PAGE):

- (i) Under reducing conditions: about 60 kD,
- (ii) Under non-reducing conditions: about 60 kD and about 120 kD;

20 (b) amino acid sequence:
includes an amino acid sequence of the Sequence ID No. 3 of the Sequence Table,

(c) affinity:
exhibits affinity to a cation exchanger and heparin, and

25 (d) heat stability:

- (i) the osteoclastogenesis-inhibitory activity is reduced when treated with heat at 70°C for 10 minutes or at 56°C for 30 minutes,
(ii) the osteoclastogenesis-inhibitory activity is lost when treated with heat at 90°C for 10 minutes,

30 the process comprising inserting a DNA including the nucleotide sequences of the sequences No. 1 and No. 2 in the Sequence Table into an expression vector, producing a vector capable of expressing a protein having the above-mentioned physicochemical characteristics and exhibiting the activity of inhibiting differentiation and/or maturation of osteoclasts, and producing this protein by a genetic engineering technique.

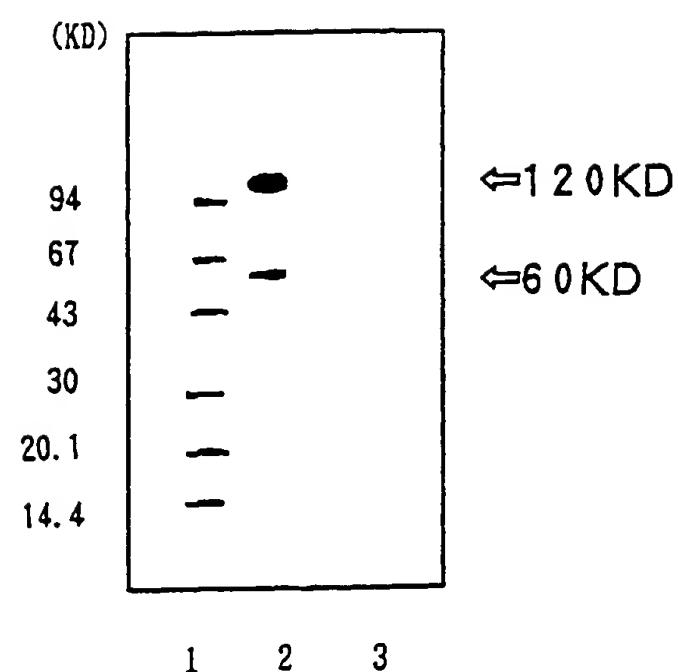
40

45

50

55

Figure 1



INTERNATIONAL SEARCH REPORT		International application No. PCT/JP97/02859									
A. CLASSIFICATION OF SUBJECT MATTER Int. Cl ⁶ C12N15/00, C12P21/00 According to International Patent Classification (IPC) or to both national classification and IPC											
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) Int. Cl ⁶ C12N15/00, C12P21/00											
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched											
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) WPI, GENETYX-CDROM, BIOSIS											
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">Category*</th> <th style="width: 80%;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="width: 10%;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>Cancer Research, (1995), Vol. 55, Toshiyuki Yoneda, et al. "Sumarin suppresses hypercalcemia and osteoclastic bone resorption in nude mice bearing a human squamous cancer" P. 1989-1993</td> <td>1 - 4</td> </tr> <tr> <td>A</td> <td>Proc. Natl. Acad. Sci. USA, (1990) Vol. 87 Kukita A. et al. "Osteoinductive factor inhibits formation of human osteoclast-like cells" P. 3023-3026</td> <td>1 - 4</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	A	Cancer Research, (1995), Vol. 55, Toshiyuki Yoneda, et al. "Sumarin suppresses hypercalcemia and osteoclastic bone resorption in nude mice bearing a human squamous cancer" P. 1989-1993	1 - 4	A	Proc. Natl. Acad. Sci. USA, (1990) Vol. 87 Kukita A. et al. "Osteoinductive factor inhibits formation of human osteoclast-like cells" P. 3023-3026	1 - 4
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.									
A	Cancer Research, (1995), Vol. 55, Toshiyuki Yoneda, et al. "Sumarin suppresses hypercalcemia and osteoclastic bone resorption in nude mice bearing a human squamous cancer" P. 1989-1993	1 - 4									
A	Proc. Natl. Acad. Sci. USA, (1990) Vol. 87 Kukita A. et al. "Osteoinductive factor inhibits formation of human osteoclast-like cells" P. 3023-3026	1 - 4									
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.											
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed											
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family											
Date of the actual completion of the international search September 29, 1997 (29. 09. 97)		Date of mailing of the international search report October 7, 1997 (07. 10. 97)									
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.		Authorized officer Telephone No.									